HANDBOOK ON PUBLIC SECTOR EFFICIENCY

Handbook on Public Sector Efficiency

Edited by

António Afonso

ISEG – Lisbon School of Economics and Management, Universidade de Lisboa, Portugal

João Tovar Jalles

ISEG – Lisbon School of Economics and Management, Universidade de Lisboa, Portugal

Ana Venâncio

ISEG – Lisbon School of Economics and Management, Universidade de Lisboa, Portugal

Edward Elgar PUBLISHING

Cheltenham, UK • Northampton, MA, USA

Published by
Edward Elgar Publishing Limited
The Lypiatts
15 Lansdown Road
Cheltenham
Glos GL50 2JA
UK

Edward Elgar Publishing, Inc.
William Pratt House
9 Dewey Court
Northampton
Massachusetts 01060
USA

A catalogue record for this book
is available from the British Library

Library of Congress Control Number: 2023930157

This book is available electronically in the **Elgar**online
Economics subject collection
http://dx.doi.org/10.4337/9781839109164

ISBN 978 1 83910 915 7 (cased)
ISBN 978 1 83910 916 4 (eBook)

Typeset by Westchester Publishing Services

Contents

PART III PUBLIC SECTOR EFFICIENCY IN THE EDUCATION, HEALTH, AND SOCIAL PROTECTION SECTORS

PART IV PUBLIC SECTOR EFFICIENCY ACROSS THE WORLD

Contributors

Antonis Adam, University of Ioannina, Hellenic Open University

António Afonso, ISEG, Universidade de Lisboa; REM/UECE; CESifo (Center for Economic Studies and Ifo Institute)

Tommaso Agasisti, Politecnico di Milano School of Management

Konstantinos Angelopoulos, Adam Smith Business School, University of Glasgow; CESifo, Munich

Özlem Aydin, International Monetary Fund, Fiscal Affairs Department, Economist

Amélie Barbier-Gauchard, Strasbourg University, BETA, France

Kéa Baret, BETA CNRS UMR 7522, University of Strasbourg, France; OECD, Paris, France

N R Bhanumurthy, Dr BR Ambedkar School of Economics University, Bengaluru, India

Boele Bonthuis, International Monetary Fund

David Coady, International Monetary Fund

Eduardo Costa, Nova School of Business and Economics

Xavier Debrun, National Bank of Belgium, European Fiscal Board, Catholic University of Louvain

Jean-Marc Fournier, International Monetary Fund

Víctor Giménez, Department of Business, Universitat Autònoma de Barcelona

Fabien Gonguet, IMF, Fiscal Affairs Department

Zhong-cheng Guan, Institutes of Science and Development, Chinese Academy of Sciences, Beijing; University of Chinese Academy of Sciences, Beijing

Luis Alberto Gutiérrez-Arango, Universidad Autónoma de Manizales (Colombia)

Samir Jahan, International Monetary Fund

João Tovar Jalles, Universidade de Lisboa, ISEG; Nova School of Business and Economics; IPAG Business School

Pantelis Kammas, Athens University of Economics and Business

Paulo Medas, International Monetary Fund

Ranjan Kumar Mohanty, Xavier Institute of Management, XIM University, Bhubaneswar, India

Rui Monteiro, International Monetary Fund, Fiscal Affairs Department, Technical Assistance Advisor

Caitlin O'Loughlin, Division of Supervision, Regulation & Credit, Federal Reserve Bank of Philadelphia

Oluwanbepelumi Esther Olanubi, Department of Economics, University of Lagos, Lagos, Nigeria

Sijuola Orioye Olanubi, The Nigerian Economic Summit Group, Lagos, Nigeria

Daniel Osorio-Barreto, Universidad Autónoma de Manizales

Pedro Pita Barros, BPI, Fundação "la Caixa" Professor of Health Economics, Nova School of Business and Economics

Diego Prior, Department of Business, Universitat Autònoma de Barcelona

Xian-tong Ren, Institutes of Science and Development, Chinese Academy of Sciences, Beijing; University of Chinese Academy of Sciences, Beijing

Isabel Rial, International Monetary Fund, Fiscal Affairs Department, Senior Economist

Biresh K. Sahoo, Xavier Institute of Management, XIM University, Bhubaneswar, India

Ludger Schuknecht, Vice President and Corporate Secretary, Asian Infrastructure Investment Bank

Gerd Schwartz, Independent advisor on fiscal policy and management (formerly International Monetary Fund, Fiscal Affairs Department)

Léopold Simar, Institut de Statistique, Biostatistique et Sciences Actuarielle, Université Catholique de Louvain

Yao-yao Song, School of Economics, Capital University of Economics and Business, Beijing

Douglas Sutherland, OECD, Paris, France.

Mouhamadou Sy, International Monetary Fund

Vito Tanzi, Honorary President, International Institute of Public Finance, Former Director of Fiscal Affairs at the IMF. Former Undersecretary for Economy and Finance of Italy. Former Professor and Chair of Economics at The American University

Michael Thöne, Managing Director/CEO of FiFo Institute for Public Economics, University of Cologne

Sofia Tsarsitalidou, University of Piraeus, University of Patras

Ana Venâncio, ISEG, Universidade de Lisboa; ADVANCE/CSG; CESifo (Center for Economic Studies and Ifo Institute)

Paul W. Wilson, Department of Economics, Clemson University

Guo-liang Yang, Institutes of Science and Development, Chinese Academy of Sciences, Beijing; University of Chinese Academy of Sciences, Beijing

Introduction to public sector efficiency

António Afonso, João Tovar Jalles, and Ana Venâncio

1. THE ISSUE OF PUBLIC SECTOR EFFICIENCY

This Handbook deals with Public Sector Efficiency (PSE), which has become increasingly relevant, notably for policy makers, in a context of constrained fiscal space and public funding sources. In fact, economic agents and voters are increasingly more educated and not only request but also require more accountability on the use of public money. Specifically, they wish to understand what type of public goods and services they obtain in exchange of the amount of taxes paid. Indeed, and quoting a former US President,

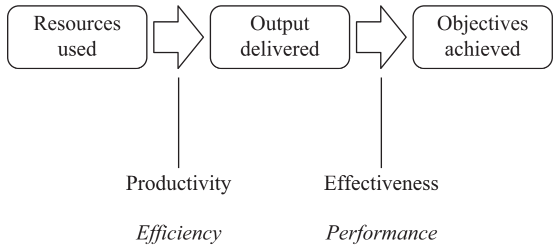
> The question we ask today is not whether our government is too big or too small, but whether it works (. . .). Where the answer is yes, we intend to move forward. Where the answer is no, programs will end. And those of us who manage the public's dollars will be held to account – to spend wisely, reform bad habits, and do our business in the light of day – because only then can we restore the vital trust between a people and their government.
>
> (Barack Obama inaugural speech, 20 January 2009)

Therefore, several main questions arise. Are 'public' services satisfactory considering the level of resources allocated? Could one have better results using the same resources? Could we have the same results with lower public expenses? Can we measure cross-country/cross-sector/cross-institution efficiency and determine benchmark units? Can we explain this measured level of public inefficiency? What factors might explain the level of public inefficiency?

Moreover, the concerns about fiscal sustainability and growing demands by taxpayers to get more value for public money as well as the need to reconsider the scope for state intervention in the economy, have prompted efforts to focus budgetary spending on more growth-enhancing and growth-inclusive activities and the allocation of resources within the public sector towards better efficiency, effectiveness and equity. This is particularly important when economies need to deal with exogenous health (such as COVID-19) or energy shocks (such as those generated from an exogenous war event such as Ukraine), which put additional pressure on strained public resources. In fact, taxpayers and policymakers (should) care about how public resources can be transformed into deliverables in terms of public goods and services. This main concern is illustrated in Figure I.1.

Additionally, when measuring public sector performance and efficiency, there is a need for good indicators and for homogenous and matching data (heterogeneity can limit the analysis), while key issues concern also mathematical and empirical methods and (homogeneous and 'right') data to assess performance and efficiency.

Regarding commonly used methods, a number of studies applied non-parametric techniques to measure public sector performance and efficiency (Herrera and Pang, 2005; Afonso, et al., 2005, 2010a, 2010b; Afonso and Aubyn, 2005, 2006; Sutherland, 2007; Adam, et al., 2011; Afonso, et al., 2013; Afonso and Kazemi, 2017; Herrera and Ouedrago, 2018). The underlying idea is simple. One or several expenditure inputs can affect one or several performance



Source: Authors.

Figure I.1 Inputs and output of public services

indicators. For instance, in the context of Data Envelopment Analysis, the most efficient countries are those on the 'frontier' of expenditure and performance. The relative distance to the frontier in terms of expenditure and outcomes shows the degree of inefficiency of the countries not on the frontier. The analysis does not argue that the countries on the frontier are in fact fully efficient. But for the lack of evidence that more efficiency is possible, it is prudent to assume that at least countries on the frontier are efficient. Such methodologies are also used notably in several of the chapters of this Handbook.

Previous studies identified substantial public sector efficiency differences between countries and the scope for spending savings in several Organisation for Economic Co-operation and Development and EU countries (see Adam et al., 2011; Dutu and Sicari, 2016; Afonso and Kazemi, 2017; and Antonelli and de Bonis, 2019), as well as in Latin American and Caribbean countries (Afonso et al., 2013). There is also evidence that capital markets reward better public sector efficiency (Afonso et al., 2022), and that tax reforms are relevant for the efficiency of government spending (Afonso et al., 2021). To better understand these cross-country efficiency differences, previous studies have examined several potential drivers, such as: population size, education, income level, the quality of institutions (property right security, corruption and so on) and also the level of quality of the country's governance, size of the government, political orientation, voter participation and civil service competence (Afonso et al., 2005; Hauner and Kyobe, 2008; Antonelli and de Bonis, 2019).

Hence, the Handbook on PSE addresses methodological and measurement issues applied to different economic questions with a policy angle in mind. The 19 chapters provide new cross-country datasets on public sector efficiency and address issues such as the importance of fiscal rules, public–private partnerships, state-owned enterprises and national and sub-national governments. It also dwells on sectoral public spending areas such as health, social protection, pensions and public investment. At the same time, several geographical regions and individual countries are covered in the efficiency analysis, such as India, China, Sub-Sahara Africa and OECD economies.

2. STRUCTURE OF THE HANDBOOK

The Handbook on PSE has four parts. More specifically, **Part I** comprises four chapters. Chapter 1 (by Afonso, Schuknecht and Tanzi) deals with the size of government. Chapter 2 (by Thone) looks at the quality of public finances. It presents also new datasets of efficiency scores for 36 OECD countries, in the period 2006–17, in Chapter 3 (by Afonso, Jalles and Venâncio) and discusses new methodologies for assessing government efficiency in Chapter 4 (by O'Loughlin, Simar and Wilson).

The six chapters in **Part II** deal with public sector efficiency and institutional quality, pensions, fiscal rules, public–private partnerships and state-owned enterprises. Specifically, Chapter 5 (by Adam and Tsarsitalidou) revisits overall public sector efficiency, Chapter 6 (by Angelopoulos and Kammas) assesses the role of historical institutional quality for public sector efficiency. Chapter 7 (by Barbier-Gauchard, Baret and Debrun) deals with the relevance of fiscal rules for government efficiency. Chapter 8 (by Bonthuis) addresses the issue of public pension spending efficiency. The fiscal risks and efficiency of public–private partnerships is the topic covered in Chapter 9 (by Schwartz, Aydin, Monteiro and Rial). Chapter 10 (by Medas and Sy) examines the main drivers of state-owned enterprises' efficiency and what governments can do to get the most out of them.

The five chapters in **Part III** address the government spending efficiency issues of secondary education, tertiary education, public health spending and social protection spending. It also deals with the issues related to the efficiency of public investment. Therefore, Chapter 11 (by Sutherland) looks at compulsory education efficiency in OECD countries. Regarding also OECD economies, Chapter 12 (by Agasisti) studies the efficiency of higher education institutions. Chapter 13 (by Barros and Costa) addresses the question of efficiency of public health spending in European countries. In Chapter 14 (by Coady and Jahan) the topic under analysis is the efficiency of social protection spending for 28 European Union countries. Chapter 15 (by Fournier and Gonguet) discuss the efficiency of public investment and implications in terms of an optimal level of public investment.

On the other hand, the five chapters in **Part IV** cover efficiency issues in different geographies. Notably regarding Colombian departments in Chapter 16 (by Gutiérrez-Arango, Giménez, Osorio-Barreto and Prior), Indian states in Chapter 17 (by Mohanty, Bhanumurthy and Sahoo), science and technology activities in Chinese Universities in Chapter 18 (by Song, Ren, Yang and Guan) and government spending in Sub-Sahara Africa in Chapter 19 (by Olanubi and Olanubi).

Finally, the Handbook will be an important reading reference for those academics, policymakers, international institutions, students and the general public, with an interest in a specific applied topic covered in one of the several chapters. Indeed, the authors of the several chapters are affiliated to academia, Central Banks, the IMF, the OECD and other international organizations.

REFERENCES

Adam, A., Delis, M., Kammas, P. (2011). 'Public sector efficiency: leveling the playing field between OECD countries', *Public Choice*, 146 (1–2), 163–183.

Afonso, A., Kazemi, M. (2017). 'Assessing Public Spending Efficiency in 20 OECD Countries', in *Inequality and Finance in Macrodynamics (Dynamic Modeling and Econometrics in Economics and Finance)*, Bökemeier, B., Greiner, A. (Eds). Springer.

Afonso, A., Jalles, J., Venâncio, A. (2021). 'Structural Tax Reforms and Public Spending Efficiency', *Open Economies Review*, 32, 1017–1061.

Afonso, A., Jalles, J., Venâncio, A. (2022). 'Do Financial Markets Reward Government Spending Efficiency?', *Journal of International Financial Markets, Institutions & Money*, 77, 101505.

Afonso, A., Romero, A., Monsalve, E. (2013). 'Public Sector Efficiency: Evidence for Latin America'. IADB Discussion Paper IDB-DP-279.

Afonso, A., St. Aubyn, M. (2005). 'Non-parametric approaches to education and health efficiency in OECD countries', *Journal of Applied Economics*, VIII (2), 227–246.
Afonso, A., St. Aubyn, M. (2006). 'Cross-country Efficiency of Secondary Education Provision: A Semi-parametric Analysis with Non-discretionary Inputs', *Economic Modelling*, 23 (3), 476–491.
Afonso, A., Schuknecht. L., Tanzi, V. (2005). 'Public sector efficiency: An international comparison', *Public Choice*, 123 (3), 321–347.
Afonso, A., Schuknecht, L., Tanzi, V. (2010a). 'Public Sector Efficiency: Evidence for New EU Member States and Emerging Markets', *Applied Economics*, 42 (17), 2147–2164.
Afonso, A., Schuknecht, L. Tanzi, V. (2010b). 'Income Distribution Determinants and Public Spending Efficiency', *Journal of Economic Inequality*, 8 (3), 367–389.
Antonelli, M., de Bonis, V. (2019). 'The efficiency of social public expenditure in European countries: a two-stage analysis', *Applied Economics*, 51 (1), 46–70.
Dutu, R., Sicari, P. (2016). 'Public Spending Efficiency in the OECD: Benchmarking Health Care, Education and General Administration'. OECD Economics Department Working Papers 1278.
Hauner, D., Kyobe, A. (2008). 'Determinants of Government Efficiency'. IMF WP/08/228.
Herrera, S., Pang, G. (2005). 'Efficiency of Public Spending in Developing Countries: An Efficiency Frontier Approach'. World Bank Policy Research Working Paper 3645.
Herrera, S., Ouedraogo, A. (2018). 'Efficiency of Public Spending in Education, Health, and Infrastructure: An International Benchmarking Exercise'. World Bank Policy Research Working Paper 8586.

PART I

SIZE AND QUALITY OF THE GOVERNMENT

1. The size of government

António Afonso, Ludger Schuknecht, and Vito Tanzi

1. INTRODUCTION

The role of government and the money it spends to fulfill its role has been one of the key questions of economics and political philosophy for centuries. However, only over the past 150 years has the spending role of the state developed significantly and, thus, has started a vivid debate on the appropriate role and size of government. And only over recent decades, when governments had grown significantly in size have methodologies emerged to measure, assess and advise on the size, performance and efficiency of government and on the underlying policies.

Government expenditure mostly takes place via the budgets of different levels of government. The sum of this spending is typically referred to as the size of government. Public spending, which we use as a synonym, comprises spending on various tasks of government. These can be categorized according to an economic classification – consumption, investment, interest, subsidies and transfers – or a functional classification – education, defense and so on (for details see OECD, 2019c).

The size of government is derived from its spending role in an economy and this is linked to the question of what governments do and should do. In this chapter, we link public spending with the role of government and describe how much governments spend and what they spend it on. We also reference non-budgetary expenditure and fiscal risks for government as these are also part of the 'size' of government in a broader sense. Whether government performs its spending role well in terms of objectives, performance and efficiency is subject to analysis and will be discussed in later chapters of this volume.

The remainder of this chapter is organized as follows. Section 2 discusses what the government should do. Section 3 addresses the issues related to how government should intervene and spend. Section 4 reviews the size of government from a historical perspective. Section 5 illustrates in a stylized way the composition of public expenditure. Section 6 concludes.

2. WHAT SHOULD GOVERNMENT DO?

In the distant past, when individuals made a living by hunting and fishing or by subsistence agriculture, there was no or little need for a government in today's sense of the word. Consequently, there was no sense in asking what a government should do. Individuals and families were largely on their own and were free to pursue their individual interests and to satisfy their needs in the best way they could. Their (very low) standards of living depended on their personal ability to collect food and to protect themselves against natural elements and other dangers. At that time, the actions of individuals were not likely to generate significant externalities that could affect, positively or negatively, other individuals,

With the passing of time individuals started to see the advantages that could come from aggregation and from operating in groups, first small groups and then progressively larger ones. They also started to make distinctions between activities and needs that could best be satisfied through their independent actions and needs that could be better satisfied though the collective action of the group. As the groups became larger and more stable (in terms of location and membership), and less homogeneous in the abilities and the attitudes of the individuals that composed them, a process of selection started to develop. The individuals in the groups, or the communities, started to be differentiated into different categories, and specialized in their jobs, to better satisfy the perceived, collective needs. Productivity increased as specialization spread, as Adam Smith in 1776 so eloquently described.

The satisfaction of a particular community needs to cater to different groups of individuals who were assumed to have greater abilities to deal with those needs. *Castes* or *classes* started developing (see Brown, 2005). Some individuals assumed the task of providing protection against dangers coming from the outside. Some (sorcerers, priests and others) were assigned the task of communicating with the gods or the divinities. Others took on the more mundane tasks of providing food and dealing with other daily needs of the community. Interaction and exchanges increasingly took place in markets, using some form of money as a medium of exchange (see, for instance, Brown, 2005, Mishkin, 2004).

The larger communities saw also the necessity, or the convenience, of having some individuals or some small groups assume the responsibility of making rules for the behavior of the individuals in the community. These were embryonic forms of government, and the assignments of the responsibilities described above were rarely made democratically. Much political science literature has shown that governments generally came into being out of the domination by one group over the rest of the community (see Loria, 1886 and Mosca, 1896). There are probably no examples in history of governments that were born as democratic and in which all individuals had equal political power and equal individual freedom.

In recent centuries, especially in the nineteenth century, several, then more advanced, countries started showing traits that could be called democratic. Some individuals were given the constitutional right to vote and, through their vote, to influence government policies. With time, this right was progressively extended to larger groups, including individuals who had no property, some of whom had been effectively *serfs* or even *slaves* in the past, and to women (see Tanzi, 2020a).This process of increasing democratization was happening at a time when markets were becoming larger in number of participants and in territories covered; and they were becoming freer from government meddling than they had been during the earlier, *mercantilist* times described by Adam Smith and by others. Democratic countries with market economies required some rules to guide the behavior of the citizens and to protect some of their rights, while restricting abuses by fellow citizens and the arbitrary power of governments.

This led to the important question of what the *economic* role of governments in countries that were democratic and that depended on free markets for the provision of goods and services needed by citizens and for the generation of incomes to those who provided the factors of production should be. What should the scope and limit of government intervention in these societies be? Although the latter question had been raised occasionally, over earlier centuries, by some philosophers and early economists, it had been raised in broader and more political contexts. In the nineteenth century this became a more important and more specific question, at a time when governments were becoming more democratic, and markets freer.

In the nineteenth century, two contrasting schools of thought competed in the *market for ideas,* and they gave very different answers to the above question. The two schools were *laissez faire* and *socialism*. Notably, Mill (1848) and several other economists argued that *laissez faire* should be the general practice, and anything else would be unwelcome, while the socialist view would postulate a stronger government presence in the economy and in society (see Engels, 1880, and the overview of Musgrave, 1985). Both of these schools became intellectually and politically important and attracted followers and influence among the common citizens and the intellectuals of the day. Generally, those who owned wealth tended to favor *laissez faire*, while many workers were attracted by the socialist ideas. This fact started to influence some policies, as for example it did the Bismarck pension reforms in Germany in the late nineteenth century, when socialism had become very influential in Germany and in other countries (see Ashley, 1904).

The *laissez faire* school, which was more unified intellectually than the socialist school, had its beginning with the work of Adam Smith in 1776, and later had come to be influenced by the Darwinian evolutionary theory. It maintained that the role of the government should be limited and should not interfere with the natural evolution of societies. A free-market economy would deliver more progress and, over the long run, it would promote growing standards of living for the whole population.

The *laissez faire* school stressed the need to protect property rights and the rights of citizens to engage in legitimate economic activities, without the need for government authorization. It paid little attention to the distribution of income or to some, obvious, failures that existed in the market, including monopolies. It stressed the importance of individual liberty and the contribution that personal initiative made to economic activities. A basic assumption was that, in a free society with a free market, anyone willing to work would be able to find a job and to earn a living income. In such a society, property owners inevitably had more political and economic power than workers. Property rights tended to receive more government protection than the rights of workers to be well paid and to work in safer jobs. Social and distributional objectives were not seen as a core role of government.

The socialist schools, of which there were several branches, some much more radical than others, were concerned with the masses, with the distribution of income, and with the status of the workers. They were critical of property rights, and much less interested in *individual* liberty. They advocated a larger economic role of the state and pushed for high public spending. Some versions of socialism (especially the Marxist version) advocated the expropriation of property and the creation of governments which, through central planning, would direct production and the distribution of income toward the satisfaction of the ‘basic needs’ of the masses.

Both sides often held extreme views. Socialists saw the market role in much less favorable eyes and many of them considered property as a ‘theft’. Just how extreme *laissez faire* had, at times, become can be seen by the reaction of the leading Italian economist at that time, Francesco Ferrara. In the 1850s he wrote that, by imposing an import duty, the US federal government had committed ‘a sin as grave as that of slavery’. Similar extreme views were expressed by other leading economists, such as F. Bastiat, Gustave de Molinari and J. B. Say in France. In Germany, a socialist economist, Lassalle, was jailed for advocating in a lecture in Berlin a progressive income tax, which would be ‘a single progressive income tax in state and community, instead of all existing taxes, especially the labor-crushing indirect tax’ (see notably Spahr, 1886).

In the second half of the nineteenth and during the twentieth century, the problems that totally free markets faced in the real world were being addressed by some economists. Monopolies, which were common and were generating enormous incomes for some individuals, had started to be regulated (Wicksell, 1896, see Musgrave and Peacock, 1958). Some rights of workers (to organize, to strike and to have their working hours limited) were recognized and limits were imposed on the age at which children could work. Safety standards in workplaces were strengthened. Pension and other insurance schemes started to be created for workers and school attendance of children became mandatory. During the Great Depression, a new important government role, to fight economic recessions, would be theorized by Keynes in *The General Theory* in 1936, and it would become a government responsibility during the second half of the twentieth century.

In the decades after World War II, the pure *laissez faire* ideology of earlier years was in retreat. There was increasing skepticism, even on the part of many orthodox economists, about the assumed efficiency of the market without corrections. In addition, there was less tolerance for the income distribution that the free market generated. Progressive income taxes became more popular. In the years after World War II, there was an intense search by economists for *market failures*, beyond the supplementary role for which public goods had been known. There was a search for ways to make the income distribution more equitable. Monopolistic competition came to be seen as influencing many markets (Robinson, 1933; Chamberlin, 1948; Musgrave, 1985).

In the late nineteenth century, the size of government was still very small, given in part to the prevailing view of its limited role (see section 4). This changed, first, over two World Wars and subsequently with the ascent of the welfare state so that, in particular, the post-1960 period witnessed fast increases in public spending and in tax levels in advanced countries. It also saw a growing use of regulations, to deal with externalities that were assumed to have negative consequences for individuals or for the environment. It was clear that the economic role of the state had changed. It had become larger in countries that were still considered market economies (Tanzi and Schuknecht, 2000).

Government intervened to deal with: (a) pure public goods and quasi-public goods; (b) various market failures; (c) negative externalities; (d) business cycles; (e) income maintenance for individuals and families unable to earn a living; and, (f) finally, it intervened to make the income distributions closer to what voting societies expected them to be. Tax levels and tax revenues' ratios over Gross Domestic Product (GDP) went dramatically up, and tax systems became more progressive and, especially in some countries, more complex. New government programs were created, some aimed at dealing with *universal* risks (illness, disability, unemployment, old age and illiteracy) others, that were *means tested*, focused on economic problems of particular individuals and families. The composition of public expenditure that in the aggregate makes up the size of government is discussed in depth in section 5.

The growing economic role of the state, which Keynesian economists had propagated, was inevitably controversial. Conservative and libertarian economists, from the Chicago School, the Austrian School and the new School of Public Choice, were highly critical of the new government role. They believed that it created inefficiencies, reduced economic growth, introduced privileges for special interests and reduced the freedom and the incentives of individuals who would come to depend on a 'nanny state'. These critics believed that a state that reduced incentives for individuals and made them more dependent on the government had reduced the vitality of economies and economic growth.

For instance, one strand of economics tends to argue that the size of government reflects social preferences in more government and supports more redistribution to correct market failures (including Wagner's law of governments producing superior goods). Other political economists have emphasized the role of institutions in shaping the size of government and pointed to political market failure in leading to governments being bigger and less efficient than they should be. The size of government is affected by voting rules (Husted and Kenny, 1997), interest group competition (Becker and Mulligan, 2003), party preferences (Bräuninger, 2005), political centralization (Fiva, 2006), the prevailing income distribution (Meltzer, 1983) and the degree of openness and globalization (Shelton, 2007; Rodrik, 1998; Potrafke, 2009; Dreher et al., 2008).

Many studies have assessed the impact that a larger economic role of the state was having on macroeconomic performance and other objectives (for surveys see the relevant chapters in this volume). As is often the case in these attempts, the *a priori* political biases of those who made the attempts often predicted the results. Conservative economists tended to find higher negative results from the higher government role, while social democrat economists tended to minimize the negative impact of that role. The bottom line is that it is difficult to conclude definitely that countries that spend more, such as various European countries, have performed less well than the countries that spend less, such as the Anglo-Saxon countries, taken as a group. Afonso and Jalles (2016) report that the detrimental effect of government size on economic activity is stronger the lower institutional quality and the positive effect of institutional quality on output increases with smaller government sizes. Often the way in which the money is spent and the way in which the taxes were collected is more important than the size of the spending and of the taxing (see Tanzi, 2020a and 2020b).

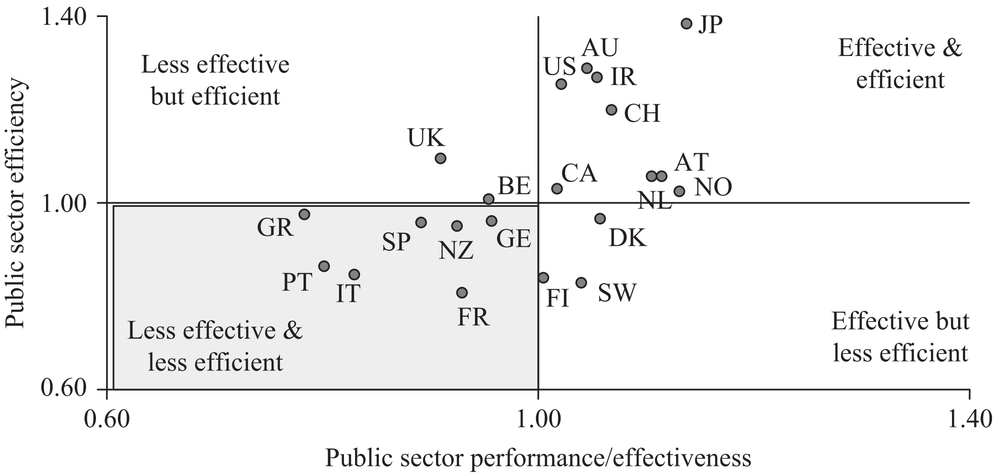
Naturally, the macroeconomic performance of the economy, though important, is just one measure of the impact of the role of government on economic welfare. In recent decades the importance of that measure has been challenged by observers who have pointed out that much of the economic growth, in several countries including the United States, has benefitted a small share of the population, while the large majority has seen little improvement in its standard of living. Therefore, there may have been *growth* but, by other measures, there may have been little genuine development or improvement of wellbeing (Hessami, 2010). For example, the recent, inequality-adjusted Human Development Index, prepared by the United Nations, lists high and low spending countries among the top performers (Davies, 2009).

It may, therefore, be possible and legitimate to focus on the impact that a larger government role has not only on economic growth but also on various socio-economic indicators that are considered important (see related chapters of this volume, also Tanzi and Schuknecht, 1997; Afonso et al., 2005; Schuknecht, 2020). Another important dimension of this theme is the interaction between the role and efficiency of the public sector and its susceptibility (or resilience) to crisis and (external) credibility for investors.

As an illustration, Figure 1.1 depicts several cases of the performance and efficiency in a European Union country sample, in 2000, where both performance and efficiency of government spending can go hand in hand.

3. HOW SHOULD GOVERNMENT INTERVENE AND SPEND?

It is very difficult, or perhaps even impossible, to provide a simple and universally appropriate list of ways in which governments should intervene in the economy, and of how much they should tax, spend and regulate. Different governments may aim at promoting different objectives, and there is no universally unquestioned way to select a list that may be optimal for all



Notes: Good performance (two right-hand side quadrants) includes lower efficiency/higher spending (Finland, Sweden and Denmark) and higher efficiency/lower spending (Austria, Japan, Ireland, US).

Source: Adapted from Afonso et al. (2005).

Figure 1.1 Illustrative evidence on public sector performance and efficiency, 2000 (considering general government spending)

societies. That list may give prominence to economic growth, without much concern for how the growth is distributed among the citizens. It could give prominence to the promotion of important socio-economic indicators, giving more weight to the importance of some of them. It could promote protection against risks and it could choose the promotion of a more equal income distribution, and other options.

In the past, economic growth received much attention. In today's democratic countries the decision on the preferred objective must be made in the democratic political process. Governments must also respect basic, fundamental, human rights, to protect minorities from possible excesses promoted by the majority, including protecting property rights, as stressed by the school of public choice. In fact, this has been a basic condition for many constitutions.

Whatever the main objective chosen is, it is important that the government's intervention not be arbitrary and that it respects some basic rules and promotes efficiently the use of scarce resources. It should not become a major drag on the economy, as it has in some countries in the past and present. The intervention should promote indicators that contribute to the economic and social welfare of the citizens. Too much emphasis on single variables, be these the growth of GDP, or changes in the Gini coefficient, is generally not desirable. Economists who suggest a single objective or a short list of objectives generally ignore the diversity of countries' situations. They thus risk stating criteria that often are not necessarily desirable or important.

Advanced countries have a different capacity to intervene, and different needs for their governments to do so, than developing countries. This difference has been recognized for a long time (see Newbery and Stern, 1987 and Tanzi, 1991 and the public finance literature related to developing countries). Countries with more even income distributions may have different reasons and justifications for the government to intervene than countries that feature more inequality. Countries with more efficient markets, in which the incomes received are considered as incomes genuinely earned and not considered rents, have more trust in the market and thus, in principle, less justification for governments to intervene, than countries with less efficient markets.[1]

Governments with public employees who respect Weberian criteria of behavior have more ability to intervene successfully than those with less efficient, politically chosen and less honest employees. When governments are less corrupt and more efficient, there is, in principle, more scope for government activity while inefficient, rent-seeking governments should be smaller (Dzhumashev, 2014). Additionally, countries that give more weight to the objectives of the collectivity (such as a more even income distribution or better management of universal risks) have more reasons for governments to intervene than countries where individual freedom is given more weight.

The first, main reason for government interventions is the provision of classic public goods. Of these, so-called pure public goods (defense, judiciary and security) are a government intervention on which most economists have agreed. However, the real-world problem with this intervention is that, while there is agreement that governments must intervene and must provide public goods, there is no guidance on how much of those goods should be provided. The theory does not provide clear guidance on what is the optimal amount of defense spending; or how much should be spent for providing protection to individuals and property; or for justice or infrastructure. The political debates in countries are focused not on whether the government should provide these public goods but on how much of them it should provide. In this, the theory is not helpful.

The debates are sharper in the provision of so-called quasi-public goods (education, health and some other provisions) for which the justification on allocation grounds combines with that on equity grounds. When a government (or private providers financed by government) is not providing good public schools, or good health services *for everyone*, it is creating different income opportunities for different categories of citizens, and it is perpetuating income differences across different categories of citizens. In this case, what has been called the ‘birth lottery’ ends up determining the future life incomes of many citizens. Countries where citizens care about avoiding large, permanent, income differences among them are more likely to want publicly financed good schools and good health services *for everyone,* to create more equal opportunities. This objective requires higher taxes and public spending.

In this context, increasing tax rates will generate deadweight burden, and the heavier the tax rates, the less they may yield relatively. The loss of utility for the individual taxpayer increases with the square of the tax (Dupuit, 1844, pp. 281). Additionally, Afonso and Gaspar (2007) illustrate numerically that financing through distortional taxation causes excess burden (deadweight loss), magnifying the costs of inefficiency.

The above arguments have implications for the level of public spending, the composition of expenditure and for the structure of the tax systems needed to finance the spending. Public programs that aim at dealing with the universal risks of all citizens, or that aim at creating more equal opportunities for more citizens, are inevitably more expensive, and must be financed by broader-based taxes requiring lower marginal tax rates than means-tested programs. The alternative of having means-tested programs, accessible by only selected groups, and accompanying them with ‘tax expenditure’, as the United States and some other countries have done, reduces the level of taxes and spending but leads to other difficulties (Tanzi, 2020a).

Finally, an issue that has been highlighted by the 2020 COVID-19 pandemic is that the traditional literature on the desirable economic role of the state has dealt with that role in equilibrium situations, and with changes at the margin of the equilibrium created by the political market. In other words, it has not dealt with shocks to the system when the role of the state may become especially important. We know that, during major wars, the government role changes and becomes particularly important. Price controls, rationing, appropriation of resources, very high marginal tax rates and other policies that are not market friendly are used. Many years ago, Peacock and Wiseman, in a classic 1961 book, also argued that wars had even changed permanently the economic role of the state.

Shocks to the economic systems of countries may come from major wars, natural catastrophes, depressions, revolutions, pandemics and other calamities. Limited government intervention may become less optimal in a world that is subjected to occasional, existential shocks. Such shocks often reveal major gaps in the role of the state. Still, it remains an open question as to whether additional resources for government would actually be spent on making countries more resilient against major shocks. The COVID-19 pandemic is a case in point: countries with very large public sectors and countries with smaller public sectors all struggled to address the issue.

With these considerations in mind, it comes as no surprise that there is no conclusive result on how big government should be and what its optimal size is. In earlier work, two of the authors of this chapter had suggested that, when the promotion of several, un-weighted, socio-economic indicators is the objective of the policy, a level of public spending of around 30–35 percent of GDP may set the desirable limit (Tanzi and Schuknecht, 2000). And earlier work by all three authors had focused on the quantitative connection that may exist between the level of public spending and some important socio-economic indicators (Afonso et al., 2005, 2010a). They found that, if these indicators reflect the desirable objectives, lower levels of public spending are possible.

As to the literature on the presumably 'optimal' size of government, lower estimates of government spending are below 20 percent for certain advanced countries (Vedder and Gallaway, 1998). Many estimates are in the 30 percent to lower 40 percent range, although optima differ very much across countries (Tanzi and Schuknecht, 2000; Pevcin, 2004; Facchini and Melki, 2013; Fort and Magazzoni, 2010; Schuknecht, 2020). Actual public expenditure is mostly significantly higher, suggesting the potential for considerable expenditure savings in many countries, even when considering the need for some spare capacities in the provision of certain public goods such as health or defense to deal with big shocks and emergencies. It could be added that some countries, including Sweden, Canada and others, in the 1990s dramatically reduced public spending while suffering no visible adverse consequences (Schuknecht and Tanzi, 2005).

4. THE SIZE OF GOVERNMENT FROM A HISTORICAL PERSPECTIVE

4.1 Public Spending Since the Late Nineteenth Century

Over the past 150 years, the size of government developed in line with the evolving thinking about its role and its capacity to raise taxes (see Tanzi, 2018). Moreover, in recent decades it reflected increasingly the growing role of social spending and the creation of the welfare state.

In the late nineteenth century, at the time when *laissez faire* was still the dominant economic philosophy, government only absorbed a very limited share of economic resources in the countries that are today's advanced economies. About 1870, when data on public finances started to be available in more countries, public expenditure averaged only 11.1 percent of GDP (Table 1.1).[2] Switzerland and Australia featured as 'big' governments, exceeding 15 percent of GDP, while Sweden and the United States reported spending well below one-tenth of GDP. This picture had not changed much before the beginning of World War I, excepting that the later warring countries, Austria, Germany and France, joined the group of relatively 'high' public spending. Public revenue in peace times was broadly in line with public expenditure, following a (mostly) unwritten rule of balanced budgets outside wars.

With World War I, public expenditure and revenue increased considerably and the protagonists of World War I reported the highest expenditure ratios: Germany, France, Italy and the United Kingdom governments spent more than one-quarter of GDP on the back of continuously high receipts after the war (Peacock and Wiseman, 1961). Just before World War II, public expenditure ratios had increased somewhat further to an average of 23.4 percent in today's advanced countries, partly in the wake of the Great Depression and partly already reflecting war preparations (Tanzi and Schuknecht, 2000). In the 90 years since 1870, the public expenditure ratio had hence roughly doubled from one-ninth to almost one-quarter of GDP.

4.2 Public Expenditure Since about 1960

In the next 60 years, public expenditure ratios doubled yet again and, yet again, there were some distinct waves. After World War II, war-related expenditure declined while some other spending increased so that the overall spending ratio increased modestly to 27.9 percent of GDP by 1960 for the reported country sample. This reflected the buildup of basic safety nets over previous decades as well as growing public services such as education and infrastructure. Some European countries reported the largest public sectors, with Austria and France reporting around 35 percent of GDP. Spain, Japan and Switzerland still featured total public expenditure below 20 percent of GDP. Revenue and expenditure had been mostly well aligned since World War II, so that, on the back of strong growth and some inflation, public debt had come down significantly across the industrialized world.

The period from 1960 to 1980 saw an unprecedented increase in public expenditure by 15 percentage points of GDP on average in just 20 years. This was the heyday of Keynesian economics when governments actively developed public services and welfare states to today's universal systems in many countries. While public expenditure averaged 43 percent in the reported country sample, it exceeded 50 percent of GDP in the small European countries of Austria, Belgium, the Netherlands, Denmark and Sweden. Another group of countries' governments, including most non-European countries, were still 'only' absorbing one-third of national resources, although this was also much above the level of 1960.

The biggest difference to earlier peacetime episodes was perhaps the growing misalignment of expenditure and revenue. By 1980 and growing thereafter, fiscal deficits had become significant and chronic. Public debts grew together with rising real interest rates and fiscal deficits. The strong increase in sovereign debt from the 1970s continued throughout the coming decades in most countries.

Table 1.1 Total expenditure by general government (% of GDP)

	About 1870	About 1913	About 1920	About 1937	1960	1980	2000	2017
Euro area								
Austria	10.5	17.0	14.7	20.6	35.7	50.0	51.0	49.1
Belgium 1/	..	13.8	22.1	21.8	30.3	54.9	49.1	52.2
Finland	..	..	..	..	..	40.0	48.0	53.7
France	12.6	17.0	27.6	29.0	34.6	46.3	51.4	56.5
Germany	10.0	14.8	25.0	34.1	32.9	46.9	44.7	43.9
Greece	..	..	..	..	..	..	46.4	48.0
Ireland 2/	..	..	18.8	25.5	28.0	48.9	30.9	26.1
Italy	13.7	17.1	30.1	31.1	30.1	40.6	46.6	48.9
Netherlands 1/	9.1	9.0	13.5	19.0	33.7	55.2	41.8	42.6
Portugal	..	..	..	..	..	32.3	42.6	45.9
Spain 1/	..	11.0	8.3	13.2	18.8	32.2	39.2	41.0
Other EU								
Denmark	..	..	..	..	..	52.7	52.7	51.9
Sweden	5.7	10.4	10.9	16.5	31.0	60.1	53.4	49.1
UK	9.4	12.7	26.2	30.0	32.2	47.6	35.4	41.1
Other advanced economies								
Australia	18.3	16.5	19.3	14.8	22.2	33.6	36.4	36.4
Canada	..	..	16.7	25.0	28.6	41.6	41.4	41.1
Japan	8.8	8.3	14.8	25.4	17.5	32.0	38.0	39.2
Korea	..	..	..	..	..	23.0	24.7	32.4
New Zealand	..	..	24.6	25.3	26.9	38.1	37.5	38.7
Singapore	..	..	..	..	..	..	19.6	..
Switzerland	16.5	14.0	17.0	24.1	17.2	32.8	33.8	34.7
US	7.3	7.5	12.1	19.7	27.0	34.9	33.7	37.8
Average 3/	11.1	13.0	18.9	23.4	27.9	43.2	42.7	43.9
Median	10.0	13.8	17.9	24.6	29.4	41.1	41.6	42.6
Standard Deviation	6.3	7.1	10.2	11.8	13.7	15.6	9.0	11.9

Notes: 1/ Central government until 1937. 2/ When taking GNP instead of GDP for Ireland, the ratios for 2000 and 2017 are 35.5% and 32.9%, respectively. 3/ Unweighted, excluding SGP (Singapore) and KOR (South Korea).

Sources: Schuknecht (2020), based on Organisation for Economic Co-operation and Development (OECD), Ameco, World Economic Outlook (WEO), Tanzi and Schuknecht (2000). Year indicated or nearest year available.

In the 1980s and 1990s, skepticism about 'big' government and a more market-friendly intellectual environment (referred to above) resulted in a major slowdown in public expenditure dynamics. By the year 2000, average expenditure ratios were not very different from 1980. A significant number of countries had undertaken expenditure reforms in the 1980s and 1990s so that expenditure ratios had declined significantly, by more than 5 percent of GDP in Belgium, Ireland, the Netherlands, Sweden and the United Kingdom. By contrast, Finland, France, Italy, Portugal, Spain and Japan experienced further increases in the public spending ratio by over 5 percent of GDP.

The 2000s were quite a rollercoaster but, on the whole, Keynesian and pro-government thinking had a major revival. Buoyant spending in the boom years of the early 2000s were succeeded by an explosion of public expenditure ratios following the financial crisis. A number of European countries undertook major expenditure savings and reforms.

The year 2017 saw total public expenditure ratios only moderately above the level of 2000 (43.9 vs 42.7 percent of GDP). However, this understates the 'true' increase in the role of government. Discounting the decline in interest spending, primary expenditure (total minus debt service) increased by 3 percent of GDP. In 2017, Belgium, Denmark, Finland and France reported the highest public expenditure ratios, above 50 percent of GDP. Most non-European countries reported a public expenditure of below 40 percent of GDP and spending in Ireland and Singapore was even below 30 percent of GDP. In several, notably large countries, including the United States, France, Japan, Italy and Spain, deficits were still significant, leaving public debt much above pre-financial crisis levels.

In this environment, the COVID-19 pandemic struck in early 2020. Estimates and projections by the International Monetary Fund (IMF) for advanced countries from the autumn of 2020 suggested a further major increase in expenditure ratios in 2020 that was expected to partially reverse with the recovery projected for 2021. Expenditure ratios were expected to increase by an average of about 9 percent of GDP to 52 percent of GDP in 2020 (Table 1.2 based on the same sample as Table 1.1). Top ratios would rise above 60 percent of GDP.

Rather than focusing on the details, it is more important to take note of the pattern: just as during the global financial crisis, expenditure ratios were expected to increase massively and very rapidly before declining again. In the global financial crisis, expenditure ratios increased between 4 percent of GDP in less affected countries and up to 11 percent of GDP in the most affected ones (Schuknecht, 2020). This also caused major increases in the public debt ratio in both episodes.

4.3 Expenditure Across Country Groups

It also interesting to look at public expenditure across country groups from a global perspective (Table 1.3).[3] Advanced countries had the highest expenditure ratios in the late 2010s. General government spending amounted to 42 percent of GDP in 2018. Amongst emerging market economies in Europe and Asia, public expenditure ratios were typically close to those prevailing in the advanced countries with smaller government sectors. Most Eastern European countries feature public expenditure between 30 and 40 percent of GDP, and Russia and China fall into the same range (Table 1.4). Note also the significant rise in public expenditure ratios during the COVID-19 pandemic, although the increase was more moderate than for the average of advanced countries.

Table 1.2 Total expenditure by general government (% of GDP)

Country	2019	2020	2021
Austria	48.2	58.3	52.5
Belgium	52.3	61.3	56.4
Denmark	49.7	57.1	53.7
Finland	53.2	59.9	57.6
France	55.6	63.1	59.1
Germany	45.2	53.9	49.0
Greece	46.2	57.3	51.3
Ireland	24.2	29.4	26.3
Italy	48.7	59.7	53.8
Netherlands	42.3	48.5	46.1
Portugal	42.7	50.5	47.3
Spain	41.9	52.7	48.0
Sweden	48.3	53.3	49.1
Switzerland	32.7	37.3	34.0
UK	38.6	53.1	45.6
Other advanced economies			
Australia	38.5	44.5	44.1
Canada	41.2	57.3	46.1
Japan	37.7	48.1	40.3
Korea	22.6	26.1	25.2
New Zealand	40.0	46.7	44.5
Singapore	14.3	28.5	16.1
US	35.7	47.2	37.4
Average 1/	43.1	52.0	47.1

Note: 1/ Unweighted average without Korea and Singapore.

Source: IMF, Fiscal Monitor, October 2020.

Table 1.3 Total expenditure by general government, by country groups (% of GDP)

	2017/18
Advanced	42.0
Emerging & Developing	32.5
G20	38.1
Africa	27.3
Latin and North America	36.6
Asia/Oceania	30.6
Europe	44.3

Sources: OECD, IMF, Schuknecht (2020).

Table 1.4 Total expenditure by general government, Eastern Europe, Russia and China (% of GDP)

	2017	2020 Est.
Czech Republic	38.9	47.4
Estonia	39.2	44.7
Latvia	33.5	42.5
Lithuania	29.7	42.0
Hungary	46.5	50.9
Poland	41.2	48.9
Slovakia	40.4	48.8
Russia	34.8	38.8
China	32.3	37.0

Sources: IMF, WEO Oct.

The average for emerging and developing countries stood at 31.3 percent of GDP in 2018 and, in a few countries, public expenditure was even below 20 percent of GDP (Annex Table A1.1).[4] Setting these numbers in perspective with the history for advanced economies shows that the size of governments in developing and emerging countries in the late 2010s was close to levels prevailing in advanced economies in the early 1960s.

Comparing continents, Europe reported the highest public expenditure ratio, near 44.3 percent of GDP, in 2018. In much of Southern and Northern Europe, this ratio was closer to 50 percent. Public expenditure in Latin and North America stood at 36.6 percent and Asia/Oceania at 30.5 percent. It is interesting to note that, amongst the large emerging economies outside Europe, Brazil reports the highest expenditure ratio at 41.6 percent, which is near the industrial country average. China at 37.6 percent, India at 27.7 percent and Indonesia at 22.2 percent show much smaller public sectors.

In Africa, spending ratios averaged 27.3 percent, ranging from over 40 percent of GDP in South Africa to well below 20 percent in Ethiopia and Nigeria. These differences reflect differences in development stages (poorer countries spend and tax less) and in the assigned role of government (Asia and America seeing less of a role for government than much of Europe). They also reflect the ability of governments to raise tax revenue (see Tanzi, 2018).

Naturally, it is not straightforward to accurately identify the effects of public sector spending on outcomes such as economic growth, and to distinguish the effect of government spending from other determinants. Moreover, comparing expenditure ratios across countries implicitly assumes that production costs for public services are proportional to GDP per capita.

4.4 Expenditure Obligations Outside the Budget

Over the past three decades, it has become increasingly evident that budgetary expenditure does not provide a complete picture of government expenditure obligations. Due to the fact that population aging, expenditure on old-age benefits and notably health, pensions and long-term care, are likely to create dynamic needs in the future, the increase in social expenditure in the coming three decades could well be several percent of GDP even under optimistic assumptions (OECD 2017a,b and 2019a,b,c; EU Commission, 2018; Schuknecht, 2020).

Moreover, financial crises (and, most recently, the COVID-19 pandemic) periodically cause sustained major expenditure increases. In the Global Financial Crisis, countries like Ireland or Greece spent more than 30 percent of GDP on bank recapitalization (IMF, 2015). Other transmission channels from the financial sector to public finances can also lead to significant costs (Schuknecht, 2019).

5. THE COMPOSITION OF PUBLIC EXPENDITURE

5.1 Public Expenditure Composition from an Economic Perspective

Total public expenditure can be decomposed in two ways. From the economic perspective, public expenditure consists of public consumption or real expenditure (broadly, spending on goods and services, wages and salaries), investment (mostly infrastructure and buildings), the service of public debt (interest payments), transfers (mostly social benefits) and subsidies (to enterprises or consumers). Most countries and international organizations such as the OECD and the IMF publish the relevant statistics and describe the underlying policies.

In the advanced countries, public consumption and transfers/social benefits are the most important expenditure components, accounting together for roughly 80 percent of spending (Table 1.5a). Wages and salaries for civil servants account for over 20 percent. Public investment is another 5–10 percent of the total.[5] Subsidies and interest expenditure are rather small components in most advanced countries.[6]

Table 1.5 Public expenditure composition, general government, 2018 or latest available year (% of GDP)

a. Economic classification

Country groups	Total spending	Public consumption	Compensation of employees	Government investment	Net interest payments	Transfers
OECD	41.5	16.6	9.9	3.3	1.2	19.7
Advanced	42.0	16.7	9.9	3.3	1.1	21.4
Emerging & Developing	31.3	16.9	8.1	3.1	2.2	9.9
Africa	27.3	NA	7.6	2.8	3.0	8.5
Latin and North America	36.6	16.8	9.5	2.6	2.5	13.2
Asia/Oceania	30.6	13.8	6.9	4.2	1.0	12.6
Europe	44.3	17.7	10.6	3.2	1.2	23.0

Note: Group averages are simple averages of the countries included for each region.

Sources: OECD, IMF, Boston Consulting Group (BCG), 2018 or latest available year.

Table 1.5 (Continued)

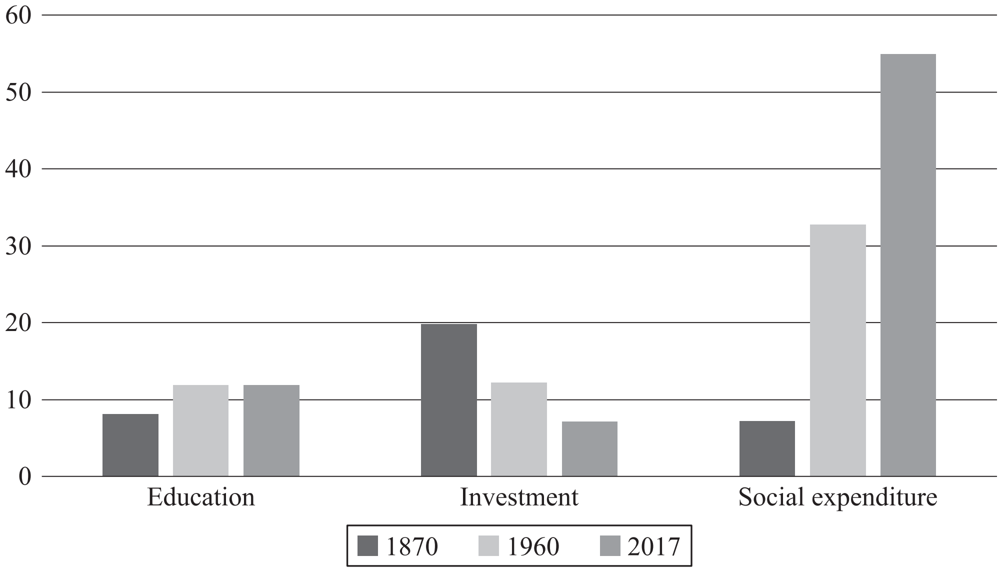
b. Functional classification

Country groups	Education	Health	Defense	Public order and safety	Environmental protection	Social Protection
OECD	5.0	6.6	1.4	1.6	0.7	15.6
Advanced	5.0	6.7	1.5	1.6	0.7	15.9
Emerging & Developing	3.4	2.3	1.2	1.6	0.2	6.0
Africa	3.6	1.5	1.2	1.5	0.2	2.9
Latin and North America	3.9	4.4	1.3	1.7	0.2	7.7
Asia/Oceania	3.7	3.8	1.6	1.4	0.6	7.0
Europe	4.9	6.4	1.3	1.6	0.7	17.8

Note: Group averages are simple averages of the countries included for each region.

Sources: OECD, IMF, BCG, 2018 or latest available year.

The expenditure composition, however, changed significantly over the past 150 years. For advanced countries, public investment as a share of total public expenditure has almost continuously declined since the late nineteenth century from about 20 percent to well below 10 percent of total spending in the 2010s (Figure 1.2). In recent decades, it also declined as a



Source: Schuknecht, 2020.

Figure 1.2 The rising share of social expenditure (% of total public expenditure) ratio of GDP. By contrast, social expenditure, which was very low 150 years ago, had increased to over half of total spending in the 2010s. Social expenditure had grown from less than 1 percent of GDP to almost one-quarter of GDP in 2016.[7]

When looking across country groups, spending on public consumption, compensation of employees, investment and even subsidies is quite similar between advanced and emerging/developing countries, even though individual country differences may, of course, be huge. Africa features somewhat higher subsidies and Asia reports above average public investment spending. Differences are significant for debt service, and Africa and (to a lesser extent) Latin America use a larger share of public expenditure to service their public debt. Advanced countries have been benefitting from near zero interest rates in the 2010s so that interest spending averaged little over 1 percent of GDP in 2018.

The biggest difference across country groups is on transfers, which comprise mainly social benefits. Advanced country spending, at 21.4 percent of GDP, is twice as high as developing and emerging country spending of 9.9 percent of GDP on this category. Europe is the biggest spender, Africa reports the lowest figures of 8.5 percent of GDP on average. Asia and the Americas report spending of 12.6 percent and 13.2 percent, respectively. In fact, the difference in the size of welfare states explains most of the differences in total spending across regions and continents.

5.2 Public Expenditure from a Functional Perspective

When looking at public expenditure from a functional perspective, there are a number of categories worth looking at in some more detail. It is interesting to note that the classic public goods – education, health, defense, and public order and safety – ‘only’ absorb a relatively modest share of public resources in all country groups (Table 1.5b).[8] Governments in advanced countries spend about 5 percent of GDP or 12 percent of their total outlays on education. The average for emerging and developing countries is somewhat lower at 3.4 percent of GDP, which is also slightly above 10 percent of total outlays. However, again, there is considerably more variation across individual countries (Table A1.1). All regions outside Europe spend slightly below 4 percent of GDP. For Africa, this is the highest share of total spending, at about 3.6 percent, which is in part due to the greater number of children of school age.[9]

Public health spending is highest in advanced countries and notably in Europe at over 6 percent of GDP. The equivalent figure is less than half of that (2.3 percent of GDP) for developing and emerging economies and only 1.5 percent of GDP in Africa. Generally, countries where citizens live longer, have higher public health spending.

Spending on external and internal security – defense and public order and safety – is quite similar across country groups. The combined spending is 3 percent of GDP, or about 7–10 percent of total spending. Environmental spending is a relatively new category, absorbing less than 1 percent of GDP in all groups. Advanced countries spend relatively much more than developing and emerging economies. Asia/Oceania and Europe spend the most but still, on average less than 1 percent of GDP. However, this is not surprising given that environmental protection is promoted mainly through regulation and taxation.

Taking these categories (excluding health) together, they amount to 10–15 percent of GDP, or about 30–35 percent of total public expenditure. This is not much, given the importance of these objectives. It illustrates that most public spending is on other things and, notably, on social expenditure.

5.3 Social Expenditure

Given its growing importance, social expenditure deserves some further discussion. Social expenditure, as defined by the OECD, includes socially related transfers and subsidies and certain other government expenditure for social purposes. The main categories are pensions, health, long-term care, family and child benefits, and unemployment. Education is mostly not included (except below primary education).

In 1960, the first year for which detailed, comparable cross-country data from the OECD became available, advanced countries spent on average 9.1 percent of GDP on social expenditure (Table 1.6). The range was enormous, from 3.5 percent in Japan to 15.4 percent in Germany. By 1980, social expenditure had almost doubled to 16.6 percent of GDP as the period after World War II witnessed the birth and expansion of many social programs. Subsequently, the ratio grew by another 2 percent of GDP per decade to 24.1 percent on average by 2016 before subsiding somewhat to 23.2 percent in the economic expansion up to 2019.

As regards individual countries, in 2019, France spent almost one-third of GDP on social matters (31 percent) followed closely by Italy and several smaller European countries. A number of other advanced countries as well as all of emerging Eastern Europe reported social spending below 20 percent of GDP (Schuknecht, 2020). In Korea, this figure was only 12.2 percent as the welfare state was still less developed. For emerging and developing countries outside Europe, comparable figures are not available but social benefits and transfers were typically lower.

The main components of social expenditure are public pensions, public health and, increasingly, long-term care. These are all related to aging of the populations and data availability is less up to date. Pension spending increased from an average of 4.5 percent in 1960 to 9.4 percent of GDP in 2014–15 in the country group of Table 1.6. The health-spending share almost tripled from 2.4 percent to 7 percent of GDP in this group. Long-term care spending was virtually inexistent in 1960 and reached several percent of GDP in a number of countries in the 2010s (OECD, 2017a and 2017b, and 2019b).

The main driver of social expenditure used to be the expansion of programs to universal coverage and to technical progress in health. In recent years, demographics have become the most important driver. It will contribute to further strong increases in the social expenditure ratio in the coming decades, if policies and benefits are not adjusted (see above).

5.4 The Financing of Public Expenditure

Government expenditure needs to be financed either from revenue, from debt, or through external grants or sales of public assets. Most expenditure is financed by domestic revenue, but the figures also show that most countries and country groups run deficits and finance part of their spending from other sources (fees, fines and so on). In 2018, the unweighted average revenue for advanced countries of 42.2 percent showed a broadly balanced budget when looking at the unweighted average (Table 1.7). However, this figure masked the fact that most large, advanced countries (including the United States, Japan and several large European countries) reported significant deficits. Emerging and development countries featured revenue of 28 percent and thus notable average deficits. For Africa, the revenue shortfall amounted to over 5 percent of GDP.

Table 1.6 Social Expenditure (% of GDP)

	1960	1980	1990	1999	2007	2009	2016	2019
Europe								
Austria	15.0	22.0	23.2	25.8	25.1	27.5	27.8	26.9
Belgium	11.4	23.1	24.4	24.6	24.9	28.6	29.0	28.9
Denmark	..	20.3	22.0	24.5	25.0	28.3	28.7	28.3
Finland	8.2	17.7	23.3	23.8	22.9	26.9	30.8	29.1
France	12.0	20.2	24.3	28.6	28.0	30.5	31.5	31.0
Germany	15.4	21.8	21.4	25.5	24.1	26.7	25.3	25.9
Greece	3.3	9.9	15.7	18.0	20.6	23.7	27.0	24.0
Ireland	7.1	15.7	16.8	13.7	15.8	22.2	16.1	13.4
Italy	10.7	17.4	20.7	22.8	24.7	27.7	28.9	28.2
Netherlands	9.6	23.3	24.0	19.1	19.9	21.6	22.0	..
Portugal	..	9.5	12.2	17.2	21.8	24.6	24.1	22.6
Spain	..	15.0	19.2	19.8	20.8	25.4	24.6	24.7
Sweden	12.6	24.8	27.2	28.0	25.5	27.7	27.1	25.5
Switzerland	4.2	12.8	12.1	17.0	16.8	18.6	19.7	15.9
UK	9.7	15.6	15.2	17.7	19.5	23.0	21.5	20.6
Other advanced economies								
Australia	5.9	10.3	13.1	17.3	15.9	17.0	19.1	16.7
Canada	8.1	13.3	17.5	16.0	16.2	18.0	17.2	18.0
Japan	3.5	10.2	11.1	16.0	18.5	21.9	23.1	22.3
South Korea	..	..	2.7	5.8	7.1	8.5	10.4	12.2
New Zealand	11.4	16.7	20.5	18.7	18.1	20.4	19.7	19.4
Singapore	..	..	..	..	..	..	..	
US	7.0	12.8	13.2	14.2	15.9	18.6	19.3	18.7
Average 1/	9.1	16.6	18.9	20.4	21.0	23.9	24.1	23.2

Note: 1/ Unweighted, excluding South Korea and Singapore.

Sources: OECD and Schuknecht (2020).

It is also worthwhile throwing a quick glance at the revenue composition. Advanced countries manage to finance over one-third of their spending or 15.2 percent of GDP with direct taxes on labor income and on profits. Indirect taxes are much more important in developing and emerging countries where they finance almost a third of all spending (9.8 percent of GDP). This figure is about 40 percent in Africa. Social security contributions are quite important in advanced countries and contribute 9 percent of GDP to total revenue. This figure is 10.8 percent in Europe but only 4.4 percent in the Americas, 2.7 percent in Europe and as low as 0.5 percent of GDP in Africa.

Table 1.7 Total revenue and revenue composition, general government, 2018 or latest available year (% of GDP)

Country groups	Total revenue	Direct taxes	Indirect taxes	Social contributions
OECD	41.0	13.9	11.7	8.8
Advanced	42.2	15.2	11.7	9.0
Emerging & Developing	28.0	7.2	9.8	3.0
Africa	21.9	7.4	8.7	0.5
Latin and North America	32.3	8.0	8.9	4.4
Asia/Oceania	28.9	8.9	9.0	2.7
Europe	44.3	14.2	12.8	10.8

Note: Group averages are simple averages of the countries included for each region.

Sources: Total revenue averages are based on data from the OECD (2018), except for Argentina, Australia, China, Colombia, Indonesia, Japan, Korea, Mexico, New Zealand, Russia, Switzerland, United States (OECD, 2017a,b), Brazil, Chile, Cote d'Ivoire, Ethiopia, Kenya, Saudi Arabia, Singapore, South Africa, Thailand, Vietnam (IMF, 2018), Egypt (IMF, 2015). Averages for taxes and social contributions are based on data from the IMF (2018), with the exception of Egypt, where data are from 2015; Mexico, India, 2017; New Zealand, 2019; data for Vietnam and India are from BCG, 2013 and 2017, respectively.

6. CONCLUSIONS

Economists today would probably all agree that governments should provide certain core public goods via public expenditure: defense and internal security, public infrastructure, public education and basic social safety nets. When looking at the size of the state that finances these goods and services, however, there are remarkable differences over time and across countries.

One hundred and fifty years ago, when the role of government started to develop in the direction of modern states, the governments of what have become today's advanced economies spent barely more than one-tenth of national income. By 1960, 60 years ago, the picture had changed completely and governments spent almost 30 percent of GDP as public goods and services and social security systems expanded. By the 2010s, advanced countries spent typically between 30 and 55 percent of GDP, with social expenditure absorbing an ever-greater share of total spending. Spending on security, infrastructure and education absorbed little more than 10 percent of GDP.

While a number of advanced countries spend not much more than 30 percent of GDP, there are also bigger governments whose social and economic performance seems to be high. Hence, there is no 'optimal' size of government, even though most if not all governments could probably become more efficient and, thus, spend less. Moreover, financial (or health) crises can result in large, sudden increases of expenditure ratios. Countries with low spending (and low public debt), by definition are likely to have more buffers to accommodate such crises without doubt about the sustainability of public finances.

By contrast, emerging and developing countries typically feature much smaller states than advanced countries and there are few exceptions to this. Some countries are fast growing and with small states (such as Vietnam or Indonesia) so that spending of 20–30 percent of GDP seems sufficient for them – similar to Western countries around 1960. Some countries, by contrast feature governments that are very 'poor' with low revenue and inefficient services contributing to little growth and development. An expenditure ratio of 10–15 percent of GDP is not enough to finance a well-functioning modern state.

Looking forward, the challenges across country groups, therefore, look very different. Advanced countries need good and efficient, not necessarily bigger, government and sometimes government may perhaps already be too big. Moreover, they must make sure that the size of government remains financeable. Core services need to be of high quality and well financed, and social expenditure or financial crisis costs must not undermine fiscal sustainability.

Emerging economies may not need to spend more but they do need to have governments adapting to the needs of their countries transitioning to more advanced economies. Increasing demands for welfare benefits and population aging are likely to raise their size of government. Developing countries often still struggle with providing well-functioning services, basic safety nets and a strong tax administration. However, more and more governments are demonstrating that progress is feasible also in that country group.

Table A1.1 Public expenditure, general government, by countries and country groups, 2018 or latest available year (% of GDP)

Countries	Total spending	Public consumption	Intermediate consumption	Compensation of employees	Government investment	Net interest payments	Transfers	Subsidies
Cote d'Ivoire	23.3	NA	NA	7.1	2.4	1.8	8.1	2.4
Egypt	32.8	NA	NA	8.2	2.1	7.3	13.4	6.1
Ethiopia	14.2	NA	NA	1.3	3.7	0.5	7.0	0.0
Kenya	28.2	NA	NA	8.8	2.2	3.7	6.4	0.7
Morocco	37.6	NA	NA	5.3	NA	1.1	13.6	6.0
Nigeria[1]	12.5	NA	NA	NA	3.3	NA	1.2	NA
South Africa	42.3	22.6	NA	14.6	3.2	3.7	10.2	0.7
Argentina[1]	41.8	NA	NA	11.7	2.0	3.8	9.3	1.4
Brazil	41.6	20.9	3.5	8.4	1.9	4.5	19.5	0.3
Canada	41.3	NA	7.3	12.5	3.9	1.7	NA	0.9
Chile	23.8	16.6	NA	8.8	2.3	0.9	12.5	1.1
Colombia	43.7	15.5	5.5	7.3	3.4	1.1	11.9	0.1
Mexico	26.3	NA	3.2	8.4	1.7	2.1	11.1	0.4
United States	38.0	14.1	6.2	9.5	3.2	3.4	14.9	0.3
Australia	36.7	17.1	7.9	9.1	3.6	1.1	14.4	1.2
China	37.1	NA	NA	6.4	5.6	0.6	NA	NA
India	27.7	NA	NA	1.4	NA	3.1	NA	NA
Indonesia	22.2	9.7	3.9	5.3	3.7	1.4	3.8	1.1
Korea	32.4	13.4	4.3	6.8	5.1	-0.5	14.0	0.3
Japan	38.7	12.7	3.7	5.4	3.9	0.3	24.1	0.5
New Zealand	38.8	16.2	6.3	8.8	4.1	0.7	16.9	0.3
Saudi Arabia	40.2	NA	NA	16.4	NA	0.5	10.7	0.5
Singapore[1]	13.9	NA	NA	3.8	NA	NA	NA	NA
Thailand	21.2	NA	NA	6.0	3.6	1.0	4.5	0.8
Vietnam	27.5	NA	NA	NA	NA	1.5	NA	NA
Austria	48.5	17.8	6.1	10.4	3.0	1.3	27.4	1.5

Table A.1 (Continued)

Countries	Total spending	Public consumption	Intermediate consumption	Compensation of employees	Government investment	Net interest payments	Transfers	Subsidies
Belgium	52.4	17.2	4.1	12.3	2.4	1.9	31.5	3.3
Denmark	51.4	NA	8.8	15.3	3.3	0.2	NA	1.7
Finland	53.1	NA	10.7	12.3	4.1	0.0	NA	1.2
France	56.1	NA	5.0	12.5	3.5	1.6	NA	2.7
Germany	43.9	13.7	4.8	7.6	2.3	0.7	28.5	0.9
Greece	46.7	NA	4.4	11.8	2.3	3.1	NA	0.8
Ireland	25.7	NA	3.4	7.0	2.1	1.6	NA	0.5
Italy	48.6	NA	5.6	9.8	2.1	3.5	NA	1.5
Netherlands	42.2	14.9	5.8	8.2	3.1	0.7	24.0	1.2
Norway	48.7	23.3	6.7	14.7	5.4	-2.4	21.3	2.1
Poland	41.5	NA	5.6	10.1	4.7	1.3	NA	0.4
Russia	37.6	15.9	5.8	10.2	4.0	0.9	15.2	0.6
Spain	41.3	17.3	5.0	10.5	2.1	2.3	21.3	1.0
Sweden	49.9	24.3	8.0	12.7	4.7	-0.1	23.9	1.6
Switzerland	34.2	13.5	4.9	7.6	3.0	0.2	17.9	3.1
Turkey	34.5	NA	5.5	8.0	NA	3.0	NA	NA
United Kingdom	40.8	18.6	8.0	9.0	2.6	2.1	19.4	0.9

Note: [1] Data for Argentina, Nigeria and Singapore are for Central Government only.

Sources: Data on total expenditure are from OECD (2018), except for Argentina, Australia, Colombia, China, Indonesia, Japan, Korea, Mexico, New Zealand, Russia, Switzerland, United States (OECD, 2018, Brazil, Chile (OECD, 2018), Cote d'Ivoire, Kenya, Nigeria, Saudi Arabia, South Africa, Singapore, Thailand (IMF, 2018), Egypt (IMF, 2015), Morocco (IMF, [illegible]), Ethiopia (BCG, 2018), Vietnam (BCG, 2013).

Data on government investment are from OECD (2018), except for Argentina, Australia, China, Chile, Colombia, India, Japan, Korea, Mexico, New Zealand, Russia, South Africa, Switzerland, United States (OECD, 2018, Côte d'Ivoire, Thailand (IMF, 2018), Egypt (IMF, 2012), Ethiopia and Nigeria (BCG, 2013), Kenya (BCG, 2012),

Data on net interest payments are from OECD (2018), except for Australia, Colombia, India, Japan, Korea, Mexico, New Zealand, South Africa, Switzerland, United States (OECD, 2018), Argentina, Chile, Cote d'Ivoire, Kenya, Russia, Thailand, Turkey, Saudi Arabia, Vietnam (IMF, 2018), China (IMF, 2018), Egypt (IMF, 2015), Ethiopia (BCG, 2018), India (BCG, 2017).

Data on transfers are from IMF (2018), except for Mexico (2017), Egypt (2015), Morocco (2011). Data are from BCG for Ethiopia (2018) and Nigeria (2013).

Data on subsidies are from OECD (2018), except for Argentina, Australia, Colombia, India, Japan, Korea, Mexico, New Zealand, Russia, Switzerland, United States (OECD, 2018), Chile, Côte d'Ivoire, Kenya, Thailand, Saudi Arabia, South Africa (IMF, 2018), Egypt (IMF, 2015), Ethiopia (BCG, 2018).

Table A1.1 (Continued)

Countries	Education	Health	Defense	Social Protection[2]	Environmental protection	Public order and safety
Cote d'Ivoire	4.9	1.3	1.5	0.2	0.3	0.9
Egypt	3.9	1.6	2.0	9.5	0.1	2.0
Ethiopia	2.1	0.6	0.6	0.7	0.1	0.3
Kenya	4.6	0.7	1.6	1.1	0.1	1.7
Morocco	NA	NA	NA	NA	NA	NA
Nigeria	0.5	0.3	0.4	0.4	NA	0.4
South Africa	5.7	4.4	0.9	5.5	0.3	3.5
Argentina[1]	1.3	1.7	0.4	1.1	0.1	1.0
Brazil[1]	2.3	2.1	0.7	14.6	0.1	1.1
Canada	NA	NA	NA	NA	NA	NA
Chile	5.2	4.0	1.0	6.2	0.2	2.0
Colombia	4.8	4.9	1.3	9.0	0.6	2.2
Mexico	NA	NA	NA	NA	NA	NA
United States	6.0	9.3	3.2	7.6	0.0	2.0
Australia	5.8	7.2	2.2	9.7	0.9	2.0
China	3.7	3.0	1.3	7.2	0.7	1.5
India	0.5	0.3	1.6	NA	NA	NA
Indonesia	2.8	1.4	0.7	1.3	0.2	1.1
Korea	5.2	4.3	2.5	6.6	0.8	1.3
Japan	3.3	7.6	0.9	16.1	1.2	1.2
New Zealand	5.8	6.8	1.0	10.8	0.9	2.0
Saudi Arabia	NA	NA	NA	NA	NA	NA
Singapore[1]	2.7	2.1	3.0	1.0	0.3	1.2
Thailand	3.1	1.2	1.2	3.0	0.0	1.2
Vietnam	NA	NA	NA	NA	NA	NA
Austria	4.8	8.2	0.6	20.5	0.4	1.4
Belgium	6.3	7.7	0.8	19.5	0.9	1.7
Denmark	6.5	8.4	1.2	22.4	0.4	0.9
Finland	5.7	7.1	1.3	24.9	0.2	1.1
France	5.4	8.0	1.8	24.3	0.9	1.6
Germany	4.1	7.1	1.0	19.4	0.6	1.5
Greece	3.9	5.2	2.5	19.4	1.3	2.1
Ireland	3.3	5.1	0.3	9.5	0.4	1.0
Italy	3.8	6.8	1.3	20.9	0.9	1.8
Netherlands	5.1	7.6	1.1	15.9	1.4	1.9

Table A1.1 (Continued)

Countries	Education	Health	Defense	Social Protection[2]	Environmental protection	Public order and safety
Norway	5.6	8.5	1.7	19.8	0.9	1.2
Poland	4.9	4.7	1.7	16.4	0.4	2.1
Russia	3.4	3.2	1.8	11.5	0.1	2.2
Spain	4.0	6.0	0.9	16.6	0.9	1.8
Sweden	6.8	6.9	1.2	20.2	0.3	1.3
Switzerland	5.6	2.2	0.8	13.5	0.6	1.7
Turkey	3.7	5.1	1.8	9.9	0.4	2.1
United Kingdom	4.6	7.4	1.9	15.2	0.7	1.8

Notes: [1] Data for Argentina, Brazil and Singapore are Central Government only.
[2] Social protection expenditure includes spending related to sickness and disability, old age, family and children, unemployment, housing, social exclusion and Research and Development (R&D) on social protection.

Sources: Data for OECD countries except New Zealand and Turkey are from OECD (2017a,b). Data are from IMF for New Zealand, Argentina, Brazil, Russia, Singapore, South Africa, Thailand and Turkey (2018), and Egypt (2015). Data are from BCG for Ethiopia and Kenya (2018), India (2017), Cote d'Ivoire (2014) and Nigeria (2013).

NOTES

1. Caution is needed with this argument, however, as additional government intervention may further worsen the functioning of markets. For the link between governments spending and inequality see Afonso et al. (2010b).
2. We always refer to figures for general government, except in a few instances of historical data. General government includes central, regional and local government and public social security as its most important components.
3. Note that regional aggregates may differ somewhat depending on the sample composition, weighting and the data source. See also OECD (2019a) for a discussion of budgeting and budget procedures and OECD (2019c) for an overview of spending categories and data. Regional spending data for 2020 were not yet available at the time of writing.
4. Note that expenditure ratios for individual countries may differ significantly across sources, especially for non-OECD countries. This, however, should not affect patterns across countries and country groups.
5. See IMF (2019) for a discussion of public investment and its management. Public investment, some argue, is always productive by definition. However, there are also studies finding the opposite with high spending correlating with much corruption (Tanzi and Davoodi, 1998).
6. There is a considerable literature on subsidies and their role in the economy. See for example Beers and de Moor (2001).
7. This risks of crowding out other, more productive spending and undermining fiscal sustainability. Schuknecht and Zemanek (2020) call this the risk of social dominance in public expenditure. Note in Figure 1.2, the country sample for 1870 is much smaller than for later years (see Table 1.1). However, the pattern across countries (of very small social spending shares) and over time is not affected.
8. See OECD (2017a, 2017b, 2018 and 2019b) for a discussion of education, pension and health expenditure.
9. For government spending efficiency assessments on education and health see, notably, Afonso and St. Aubyn (2006, 2011).

REFERENCES

Afonso, A., Gaspar, V. (2007). ‘Dupuit, Pigou and cost of inefficiency in public services provision’, *Public Choice*, 132 (3–4), 485–502.

Afonso, A., Jalles, J. (2016). ‘Economic performance, government size, and institutional quality’, *Empirica*, 43 (1), 83–109.

Afonso, A., Schuknecht, L., Tanzi, V. (2005). ‘Public sector efficiency: an international comparison’, *Public Choice*, 123, 321–347.

Afonso, A., Schuknecht, L., Tanzi, V. (2010a). ‘Public sector efficiency: evidence for new EU members and emerging markets’, *Applied Economics*, 42 (17), 2147–2164.

Afonso, A., Schuknecht, L., Tanzi, V. (2010b). ‘Income distribution determinants and public spending efficiency’, *Journal of Economic Inequality*, 8 (3), 367–389.

Afonso, A., St. Aubyn, M. (2006). ‘Cross-country efficiency of secondary education provision: a semi-parametric analysis with non-discretionary inputs’, *Economic Modelling*, 2006, 22 (3), 476–491.

Afonso, A., St. Aubyn, M. (2011). ‘Assessing health efficiency across countries with a two-step and bootstrap analysis’, *Applied Economics Letters*, 18 (15), 1427–1430.

Ashley, W. (1904). *The Progress of the German Working Class in the Last Quarter of a Century.* London, New York and Bombey: Longman, Green and Co.

Becker, G., Mulligan, C. (2003). ‘Deadweight costs and the size of government’, *Journal of Law and Economics*, 46(2), 293–340.

Beers, C. van, de Moor, A. (2001). *Public Subsidies and Policy Failures*. New York: Edward Elgar.

Bräuninger, T. (2005). ‘A partisan model of government expenditure’, *Public Choice,* 125(3–4), 409–429.

Brown, R. (2005). *Marco Polo: The Incredible Journey*. Phoenix Hill, UK; Sutton Publishing Limited.

Chamberlin, E. (1948). *The Theory of Monopolistic Competition*, 5th Edition. Boston: Harvard University Press.

Davies, A. (2009). ‘Human development and the optimal size of government’, *Journal of Socio-Economics*, 38(2), 326–330.

Dreher, A., Sturm, J., Ursprung, H. (2008). ‘The impact of globalization on the composition of government expenditures: evidence from panel data’, *Public Choice*, 134, 263–292.

Dupuit, J. (1844). ‘De la Measure de l’Utilité des Travaux Publiques’, *Annales des Ponts et Chaussées.* 8. Translated and reprinted in: Kenneth Arrow and Tibor Scitovski (eds.), 1969, AEA Readings in Welfare Economics, AEA: 255–283.

Dzhumashev, R. (2014). ‘The two-way relationship between government spending and corruption and its effects on economic growth’, *Contemporary Economic Policy*, 32(2), 403–419.

Engels, F. (1880). *Socialism: Utopian and Scientific*. Marx/Engels Selected Works, Volume 3, pp. 95–151. Progress Publishers.

European Commission (2018). ‘Ageing Report: Economic and Budgetary Projections’. Brussels: European Commission Institutional Paper 079.

Facchini, F., Melki, M. (2013). ‘Efficient government size: France in the 20th century’, *European Journal of Political Economy*, 31, 1–14.

Fiva, J. (2006). ‘New evidence on the effect of fiscal decentralization on the size and composition of government spending’, *FinanzArchiv/Public Finance Analysis*, 62 (2), 250–280.

Forte, F., Magazzino, C. (2010). ‘Optimal Size of Government and Economic Growth in EU-27’. Working Papers 0410, CREI Università degli Studi Roma Tre.

Hessami, Z. (2010). ‘The size and composition of government spending in Europe and its impact on well-being’, *Kyklos*, 63 (3), 346–382.

Husted, T., Kenny, L. (1997). ‘The effect of the expansion of the voting franchise on the size of government’, *Journal of Political Economy*, 105 (1), 54–82.

International Monetary Fund (2015). *Fiscal Monitor*. Washington, DC.

International Monetary Fund (2018). *Public Investment Management Assessment – Review and Update*. Washington, DC.

International Monetary Fund (2019). https://www.imf.org/external/np/fad/publicinvestment/pdf/PIMA.pdf

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. London: Macmillan.

Loria, A. (1886/1913). *Le Basi Economiche Della Costituzione Sociale*. Milano, Torino, Roma: Fratelli Bocca Editori.

Meltzer, A., Richard, S. (1983). ‘Tests of a rational theory of the size of government’, *Public Choice*, 41 (3), 403–418.

Mishkin, F. (2004). *The economics of money, banking, and financial markets*, 7th ed. Pearson, Addison Wesley.

Mill, J. (1848). Principles of political economy, reprinted in: 1965, Reprints of economic classics. New York: A.M. Kelley.

Mosca, G. (1896/1938). *Elementi di Scienza Politica.* Roma: Fratelli Bocca.

Musgrave, R., Peacock, P. (eds.). (1958). *Classics in the theory of public finance*. London: Macmillan.

Musgrave, R. (1985). 'A Brief History of Public Doctrine', in Auerbach, A. and Feldstein, M. (eds.) *Handbook of Public Economics*, vol. 1. North-Holland: Elsevier Science Publishers B. V.
Newbery, D. M. G., Stern, N. H. (eds) (1987) *The Theory of Taxation for Developing Countries*. Oxford: Oxford University Press.
OECD (2017a). *Health at a Glance*. Paris.
OECD (2017b). *Pensions at a Glance*. Paris.
OECD (2018). *Education at a Glance*. Paris.
OECD (2019a). *Budgeting Outlook*. Paris.
OECD (2019b). *OECD Global Pension Statistics*. Paris.
OECD (2019c). *Government at a Glance*. Paris.
Peacock, A., Wiseman, J. (1961). *The Growth of Public Expenditure in the United Kingdom.* Princeton University Press.
Pevcin, P. (2004). 'Economic output and the optimal size of government', *Economic and Business Review for Central and South-Eastern Europe*, 6 (3), 213–227.
Potrafke, N. (2009). 'Did globalization restrict partisan politics? Empirical evaluation of social expenditures in a panel of OECD countries', *Public Choice*, 140 (1–2), 105–124.
Robinson, J. (1933). *Economics of Imperfect Competition*. London: MacMillan and Co.
Rodrik, D. (1998). 'Why do more open economies have bigger governments?' *Journal of Political Economy*, 106 (5), 997–1032.
Schuknecht, L. (2019). 'Fiscal–Financial Vulnerabilities'. Frankfurt: SAFE White Paper 62.
Schuknecht, L. (2020). *Public Spending and the Role of the State*. Cambridge: Cambridge University Press.
Schuknecht, L., Tanzi, V. (2005). *Reforming Public Spending: Great Gain, Little Pain.* London: Politeia.
Schuknecht, L., Zemanek, H. (2021). 'Public expenditures and the risk of social dominance', *Public Choice*, forthcoming.
Shelton, C. (2007). 'The size and composition of government expenditure', *Journal of Public Economics*, 91 (11–12), 2230–2260.
Smith, A. (1776). *The Wealth of Nations.* Many printings.
Spahr, C. (1886). 'The Taxation of Labor', *Political Science Quarterly*, 1 (3), 400–436.
Tanzi, V. (1991). *Public Finance in Developing Countries* (Edward Elgar). Spanish edition by Instituto de Estudio Fiscales (Madrid, 1995).
Tanzi, V., Davoodi, H. (1998). *The Welfare State, Public Investment, and Growth*. Springer.
Tanzi, V. (2018). *The Ecology of Tax Systems: Factors that Shape the Demand and Supply of Taxes*. Edward Elgar Publishing.
Tanzi, V. (2020a). *The Economics of Government: Complexity and the Practice of Public Finance.* Oxford University Press.
Tanzi, V. (2020b). *Advanced Introduction to Public Finance.* Edward Elgar Publishing.
Tanzi, V., Schuknecht, L. (1997). 'Reconsidering the Fiscal Role of Government: The International Perspective', *American Economic Review*, 87 (2), 164–168.
Tanzi, V., Schuknecht, L. (2000). *Public Spending in the 20th Century. A Global Perspective*. Cambridge: Cambridge University Press.
Wicksell, K. (1896). *Finanztheoretische Untersuchungen und das Steuerwesen Schwedens*. Fisher, Jena.

2. The quality of public finances

Michael Thöne

1. THE TWO (AND MORE) DIMENSIONS OF THE QUALITY OF PUBLIC FINANCES

At the turn of the millennium, nobody was familiar with the concept of 'quality of public finances' (QPF) because the term had not yet found its way into the public finance discussion. Less than ten years later, the concept of quality of public finances covered such a broad field that – conversely – the abundance of issues under consideration made it difficult to know what exactly was meant by 'QPF'.

This broad spectrum of terminology has remained intact for another decade into the 2020s. At the core of the concept, the quality of public finances focuses on the effects that public spending – particularly in its composition – has on long-term macroeconomic targets such as economic growth and productivity progress. This chapter concentrates on this economic root of the QPF concept.

It can be argued that this focus is only one aspect of what the concept 'quality of public finances' covers as a whole. This is certainly true. For one thing, the *expenditure* side would also have to be contrasted with the *revenue* side of the public budget. Tax systems, in their composition and specific design, can significantly affect innovation and economic growth. Nevertheless, the revenue side has not gained much importance within the QPF concept. Not because of its low importance, but conversely because of its already high relevance: Public economics and actual tax policy in most countries have been keeping a close eye on this interaction for many decades. Here, the QPF concept was not needed to raise awareness of the issues.

In addition to the revenue side, which would be included in a comprehensive understanding of the quality of public finances, the discussion also covers the topic of this handbook − the efficiency of public action and public spending. The QPF concept in a broad sense considers two closely related but distinct questions in public spending: *Where* to spend it and *how* to spend it. *Composition* and *efficiency*. Is the ideal case of high-quality fiscal policy achieved when a large and growing share of public expenditure is spent on future-oriented purposes and this in an efficient way that promises high value for money?

Yes and no. Yes, insofar as these two dimensions of the quality of public finances are of course cumulative in political terms. To withstand the inherent tendency towards ever-increasing consumptive spending seems like a good thing. Spending this money wisely in the most cost-effective way attainable looks an even better thing.

Yet on the *conceptual* level, composition and efficiency of public spending should be treated as two distinctively different dimensions of QPF. Obviously, the claim to spend public money in the most efficient way attainable must not be restricted to growth-enhancing, forward-looking expenditure. The efficiency dimension of QPF covers *all* areas of the public budget. On the other hand, the composition of public spending may, as will be shown in this chapter, influence the outlooks of an economy even without regard to the bottom-up perspective of how cost effectively the money is spent in each case.

These two main dimensions of the quality of public finances – composition and efficiency of spending – have been part of the concept from very early days. This becomes clear from the comprehensive documentation of the group that – although not necessarily the birthplace – was certainly the 'nursery' of the QPF concept: The 'Working Group on the Quality of Public Finances' (WGQPF; 200–07) of the Economic Policy Committee of the EU defined the boundaries of the concept, which are still valid today (see Deroose and Kastrop 2008). Recent work emphasizes the multidimensionality of the concept (see Rodriguez-Vives 2019).

Nonetheless, this chapter only looks at one, the more or less original dimension of the QPF concept, that is, the composition of public finances. I do this in the full understanding that this handbook nevertheless covers both dimensions in their entirety.

Moreover, the limitation to 'QPF in the stricter sense' is, to a certain extent, a reaction to the multidimensionality outlined above. For this is both a blessing and a curse. It encompasses a very broad spectrum of what is really important for good fiscal policy. The width of this spectrum, however, also results in the difficulty that no one really knows what is meant when we talk about the quality of public finances. Recalling that the term was at the outset primarily used to refer to the *composition* of public spending can help to gain a little more conceptual clarity without denying the other dimensions their (great) importance.

2. FROM QUANTITY TO QUALITY OF PUBLIC FINANCES: EXPANDING THE SCOPE OF FISCAL GOVERNANCE

Not everything that is good is an investment. And not all things made of concrete have a long-term benefit. These insights – related to public spending – are older in general economics than in public economics and fiscal policy. Here, and to some extent in the public perception, for a long time government investment was virtually synonymous with forward-looking, provident policy. Conversely, government consumption expenditures, which are defined as the (extensive) residual of all non-investment expenditures, have the reputation of being short-sighted, opportunistic and wasteful. However, the dualism of investment and consumption expenditures in budgetary law and, similarly, in national accounting is unsatisfactory from both sides when it comes to the actual, empirically observed effects of public spending. For sure, not every budget item recorded as public investment in the statistics can be considered *productive* investment in the economic sense. Conversely, some types of consumptive spending provide important economic benefits in the form of provisions for the future.

At the turn of the millennium, this perspective gave rise to a discussion among public finance experts in the European Union on a different and new understanding of public investment and the concept of capital that is relevant for government activities. The background to this discussion was the quantitative consolidation of public finances, which was considered imperative in many member states at the time to comply with the Stability and Growth Pact in the long term. In this context, qualitative aspects of public spending policy were not to be ignored. Otherwise, there was the perceived risk that quantitative consolidation steps – in a misconception of equal treatment or because of their relatively weak political 'resistance' – would hit those areas of spending particularly hard which have a positive impact on growth and sustainability. With this in mind, the European Council decided with the 'Lisbon Strategy' in 2000, among other things, to take greater account of qualitative elements in the member states' deficit reduction measures. This impetus created the basis for a political and scientific discussion of the concept of 'quality of public finances', which was launched at the time.

An important component of the QPF concept is the long-term impact of the composition of public spending. Behind this stands an expanded notion of public capital, based on empirical research.

3. ELEMENTS OF A MORE ECONOMIC UNDERSTANDING OF PUBLIC CAPITAL

Which government activities and the associated expenditures are captured in an expanded concept of capital is first a normative, then an empirical question. The capital concept is normative insofar as the future effects of government activities considered here are selective. The QPF concept focuses on those government expenditures that can be expected to increase economic productivity and thus raise the potential for economic growth. In addition, all public measures with a positive impact on natural capital in the broader sense − including environmental protection, climate protection and climate adaptation − are to be included.

Already these two objectives of future orientation do not necessarily go hand in hand. The debate on 'post-growth' actually assumes the opposite. Whether or not growth and environmental sustainability can harmonize via 'decoupling', and if so, to what extent, cannot be judged in the abstract. That, too, is an empirical question that will be answered in the decades to come. In any case, decoupling resource consumption and economic growth is one of the greatest and most pressing challenges of the near future.

Indeed, very good reasons speak for the choice of these two central future tasks. In particular, economic performance and ecological sustainability on the path to climate neutrality are strategic prerequisites for being able to achieve further future goals. Still, restricting the focus to these two dimensions remains a *normative* choice. Usual sustainability agendas, for example, include all 17 target dimensions of the UN Sustainable Development Goals. If one builds indicators that also cover the numerous dimensions of social sustainability, the outcome will look quite different and much more comprehensive.

In other respects, too, 'future investments' are not the only state activities that are important for a country's economic welfare. Before these, come the fundamental public goods. In a market economy, these include above all: The rule of law and the guarantee of human rights, internal and external security, the guarantee of clear property rights, the political stability of democracy, trade policy, price stability and competition policy. Nothing works without these public services; providing them is the top priority of any government. However, once they are well provided for, further spending on these purposes quickly becomes unproductive. Technically speaking, after an early optimum, their marginal economic productivity quickly declines with further expansion of spending (EU-COM 2003).

This overview of growth- and sustainability-enhancing spending reviews the empirical research literature. It builds on existing meta-analyses (see Afonso et al. 2005; Thöne 2005; Thöne and Krehl 2016; Cepparulo and Mourre 2020; Heinemann et al. 2021; Zouhar et al. 2021). The identification of forward-looking expenditures is accomplished by means of a non-formalized meta-analysis of the research literature. The result of such a meta-analysis is an overview of the government spending that can be expected to have positive growth effects and/or sustainability effects in an industrial economy.

The studies reviewed show that many consumer expenditures can also provide important benefits for the future. Above all, the high economic significance of human capital investments, especially in the education sector, is confirmed. Here, however, as with other growth- and sustainability-enhancing expenditures, the restriction applies that the positive effects are primarily found for government activities in this field, not for spending. It is immediately obvious that more spending in a promising field does not automatically mean higher performance. Accordingly, input data − that is, budget figures − must be interpreted with caution when it comes to outcomes such as, for example, the quality of school education.

3.1 Infrastructure Investments

Not every investment expenditure in the sense of budgetary law − in Germany, this would be every expenditure on physical assets of more than 5000 euros − can also be considered an economically productive investment measure. Nevertheless, the rate of accumulation of physical capital is one of the key determinants of macroeconomic growth. Empirical work regularly confirms the growth relevance of public investment; negative findings are rather rare. The initial marginal productivity of investment is very high because of its input character. This means that not all, but very many, public investments are actually very productive. Public investment crowds-in private investment and boosts economic growth (Afonso and Alegre 2011).

Public spending on infrastructure can provide a crucial input for private economic production. In general, economic theory distinguishes several channels through which infrastructure can have a positive impact on economic growth (see EU-COM 2014). First, energy, transport and other network infrastructures are intermediate inputs for firms and thus have a significant impact on their costs and ultimately on competitiveness from an international and national perspective (Pradhan and Bagchi 2013). Moreover, government investment increases aggregate demand for construction and maintenance activities (Wang 2002; Esfahani and Ramirez 2003; Phang 2003; Short and Kopp 2005; Pradhan and Bagchi 2013). Finally, government investment can provide important signals that direct private investment into key sectors of the economy (Fedderke and Garlick 2008).

In view of the situation in many mature industrialized countries, early on Kalaitzidakis and Kalyvitis (2005) pointed in a critical direction: In their empirical study, they explicitly distinguish between the growth effects of new infrastructure investments and the effects of maintenance and repair of existing infrastructures. In their study of Canada, they show that, for a well-endowed economy, the maintenance and modernization of existing infrastructure is the crucial determinant of growth.

For EU economies, the relationship between infrastructure and human capital investment is also of strategic importance. An exemplary study here was conducted by Quirino, Macas Nunes and De Matos (2014). In their study of Portugal, they conclude that public infrastructure investment contributes to economic development, especially in low-growth regions, whereas in high-growth regions, human capital investment is the main driver of productivity. In an early survey, the European Commission also points to a ‘certain consensus’ that public investment in developed industrial societies is of secondary importance compared with spending on human capital, because above a certain level of development the marginal productivity of public physical capital increasingly falls, whereas the importance of knowledge-based skills increases (see EU-COM 2003: 106 and 110).

Despite − or because of − these clear findings, it is important to emphasize what might 'go without saying': Even a *generalized* positive assessment for a specific area of spending does not, of course, equate to a positive assessment of every *individual* measure in this area. As Bertenrath, Thöne and Walther (2006) show for the example of transport investment, such empirical 'top down' assessments cannot be readily disaggregated. This means that cost−benefit analyses are required for individual projects even where growth effects are warranted at the general level.

3.2 Technical Knowledge

The growth of technical knowledge − technological progress − is a central determinant of productive economies. In industrialized countries in particular, research and development activities (R&D) are regarded as a key to high and sustained economic growth. At the same time, technological progress is a complex economic and social process in which few clear causalities can be identified. In this context, the public sector may seek to promote the accumulation of technical knowledge and technological progress by investing in research institutions and universities.

This fundamental role is also reflected in the revaluation of investment in national accounts. In Germany's national accounts, for example, all expenditure on research and development and related personnel resources have been included in fixed capital formation since 2014 (with retroactive effect from 1991). This is the implementation of the European System of Accounts (ESA) 2010 and the underlying System of National Accounts (SNA) 2008 of the United Nations. Most other OECD countries are taking a similar approach. Since then, R&D is no longer recorded as current expenditure for production purposes, but as an investment asset that is used for production purposes over several time periods. This conceptual change applies to private companies in the same way as it does to government units and non-profit organizations.

When looking at the empirical literature, the growth impact of public and private R&D investments should be assessed separately, as it is primarily private R&D investment that is perceived to have a particularly strong impact on growth. Some researchers recommend promoting the innovative strength of companies in particular with the help of tax credits, grants, patent protection and the like.

The link between public and private R&D spending has long been regarded as a particularly delicate one, with some authors arguing that the two activities complement each other, while others have shown that public R&D crowds out private activities. Only in the case of complementarity, however, can government or government-funded R&D be seen as having a positive effect on growth, because it penetrates areas where private R&D is not pursued due to positive externalities, or because public R&D provides a locomotive function and stimulates further private R&D. A review of the empirical literature reveals a majority of studies pointing to complementarity (cf. for example Diamond 1999; Guellec and van Pottelsberghe de la Potterie 2000). But at the same time, the question is never ultimately answered.

3.3 Human Capital

Human capital with its many nuances has long been at the centre of the discussion on productivity- and future-oriented public spending. On the side of academic economics, the emergence of the 'New Growth Theory' according to Uzawa (1965), Lucas (1988) and Romer (1990) has led to a breakthrough, explaining technical progress as endogenous to the 'human factor'.

Investments in this human factor can increase productivity through many channels and can therefore have a positive effect on growth. Today, many areas of family and equality policy are also included. Finally, recent empirical research has turned to participatory, inclusive growth via human capital investment and its interpersonal distribution.

The fact that education is an area of high growth relevance within the sphere of influence of the state is an obvious fact in the OECD countries and most other economies, but by no means a matter of course. Education as an individual investment in human capital yields returns of between five and 15 per cent per annum when rewarded according to the marginal product of labour (Mincer 1974). Mincer's calculations are confirmed in numerous recent reviews (see, for example, Barro and Sala-i-Martin 1995; Temple 2000; Colombier 2011; Barro 2013). In particular, the average duration – and thus the intensity – of schooling in secondary and tertiary education has a positive effect on economic growth.

Using Germany as an example, Wößmann and Piopiunik (2009) estimated the long-term costs of inadequate education in 2009 at 2.8 trillion euros over a period of 80 years. They define inadequate education as the lack of the minimum level of basic skills required for promising participation in working life. Inadequately educated students cannot exploit their full potential in the labour market. The economic growth lost here forms the basis for calculating the follow-up costs.

At the same time, a consensus has emerged in the empirical literature that the quality and dissemination of early childhood education in particular is very important both for the productivity of an economy and for its inclusive quality (see Heckman and Cunha 2007; Fritschi and Oesch 2008; Delalibera and Ferreira 2019).

Healthcare spending has been back in the spotlight of fiscal policy discussions since the beginning of the 2020 COVID-19 pandemic. From a human capital perspective, they hold a long-established place in this discussion (Bloom et al. 2019). From a human capital perspective, health is an essential growth factor because healthy workers are physically and mentally more robust, they are more productive and earn higher wages. At the same time, they are less likely to be absent from work due to illness or family disease (Bloom, Canning and Sevilla 2001). At the microeconomic level, these connections have been extensively documented (see Strauss and Thomas 1998 for an overview). Empirical evidence also shows significant positive effects of healthcare spending on overall economic growth. The importance of health for economic prosperity becomes particularly clear when one considers the impact of the past introduction of public healthcare systems in Europe, which had a significant effect on infant mortality and death rates. The associated improvement in health status in turn had a positive impact on per capita income (Strittmatter and Stunde 2013).

More important, however, is the question whether the causality in industrialized countries might be inverted. In this case, the increase in income associated with economic growth would enable people to spend more money on their health. Economic growth would become the explanatory variable of health. A hint of a possible bidirectional causality between health spending and growth is provided by Öztürk and Altun Ada (2013). However, this alone is not a sufficient reason not to include health expenditures among the growth-related expenditures.

Family policy is very ambivalent as a growth policy because its instruments are so heterogeneous. Possible positive growth effects of family policy measures can be conveyed through two channels: On the one hand, measures to improve the work–family balance result in a higher labour supply, especially from mothers with childcare-age children. Second, in an ageing society such as Germany's, measures to increase fertility expand the labour supply, which is otherwise becoming increasingly scarce, and thus also the available stock of human capital in the long term (see, for example, Alders and Broer 2004; Weil 2006). A group of researchers led by Prognos AG conducted a major evaluation of family policy benefits in Germany in the years preceding 2014.The results showed that in the difficult decision-making process of realizing the wish to have children, public family policy measures only play the role of one influence among many. The impact of family policy measures on the labour supply of mothers is different; here, public financing of childcare in particular is shown to be a very effective measure for labour supply and higher productivity.

Gender policy – if simplified to gender relations alone – is directed at all persons of all gender identities. For developing countries, significant positive growth effects can be demonstrated in the case of increased spending on health and education for women, who are often strongly disadvantaged compared to men in these countries (cf. Dollar and Gatti 1999; Agénor and Canuto 2012). Agénor and Canuto (2014) show positive growth effects of a reduction in gender inequalities for Brazil as well.

With regard to industrialized countries, Lorgelly (2000) finds contradictory evidence on productivity in her meta-analysis. In their meta-study of the empirical literature, Kabeer and Natali (2013) also point to the asymmetric nature of the relationship between gender equity and economic growth: While there is growing evidence that gender equality positively influences economic growth, the findings for the opposite direction – the influence of growth on gender relations – are very mixed.

A paper by the International Monetary Fund (2013) discusses policies on gender equality primarily in the fields mentioned here in the section on family policy, because in Western industrialized countries the rate of female labour market participation is primarily a question of the compatibility of work and family.

Overall, empirical studies on the impact of gender indicators and policies on growth are primarily available where gender equality goals coincide with other areas of growth policy, as usually the same mechanisms of impact are considered (see for example IMF 2013). These include, for example, maternity protection and government support for childcare (see above). Furthermore, programs aimed at equitable integration in the labour market ultimately have a positive impact on economic growth through this channel. The same applies to government activities that support gender equality in education and gender-equitable access to health.

3.4 Natural Capital and Climate Protection

The 'pure' objective of the quality concept, economic growth, must be questioned. In a world of finite resources and climate change, simple economic growth is no longer a goal that can be pursued without reservations. Yet, right from the outset, the QPF concept hast reflected developments towards greener growth or, better, towards a full decoupling of growth on the one side and the use of non-renewable resources and the emission of greenhouse gases on other side (Thöne 2005).

Yet, in a time of massively increasing climate protection efforts and intensive work of numerous economic sub-disciplines, at first glance it may seem surprising that there is little broad-based empirical research on the ecological sustainability effects of government spending policies. Of course, environmental spending programs are evaluated in detail for their effects. But broad studies on the direction of impact on growth are largely lacking while the opposite direction – that is, the effects of climate change on economic growth – receives good attention (see for example OECD 2021).

This paradox – irritating and somewhat annoying at first glance – is cleared up on closer reflection. Environmental policy is not spending policy. On the contrary, government spending plays only a minor role in environmental economics compared to instruments that are designed in accordance with the polluter-pays principle. These are regulatory or revenue-based instruments. This, however, does not mean that environmental policy expenditures are detrimental to environmentally sustainable development.

Since this distinction is primarily normative, it would be inadequate to base a discriminating assessment on it, given the lack of empirical evidence.

Nonetheless, expenditures that fall under environmental protection and nature conservation count here globally as effective for the benefit of natural capital. Why is that the case? Because times have changed and an up-to-date QPF concept must make amendments for that. The first contribution to the sixth report of the Intergovernmental Panel on Climate Change (IPCC 2021) is very clear about the need to intensify the efforts of climate protection in every country of the world dramatically and immediately if the "Significantly under 2.0°C" goal of the Paris Agreement adopted at the 21st Conference of the Parties of the United Nations Framework Convention on Climate Change (COP 21) in 2015 shall still be achieved.

In view of this imperative, the questions raised about instrumental optimization of climate protection lose their thrust. In the foreseeable future, the question of revenue side or expenditure side or regulatory instruments will no longer be an issue. The principle will probably have to turn much more in the direction of 'whatever helps, helps'. From such a perspective, government expenditures in favour of climate and environmental protection can almost always be regarded as having an impact on sustainability.

4. MAKING USE OF QUALITY INDICATORS IN FISCAL GOVERNANCE

Based on the evidence of the growth and sustainability effects of public spending, specific indicators can be designed to reflect this dimension of the impact of either a specific public budget or of the general government budget. Such indicators can be useful for two different purposes. Firstly, they help to better analyse the future orientation of a specific fiscal policy. From a national perspective, they can help to bolster future oriented spending or to reduce political pressure on such spending as part of a formative fiscal policy.

Secondly, quality indicators for the composition of government spending can also be used in the supranational surveillance of the several states, as the European Union monitors its member states with the help of the Stability and Growth Pact. As mentioned, it was in this context, specifically through the Lisbon Strategy of 2000, that the term 'quality of public finances' was originally coined.

For both purposes, quality indicators are helpful both longitudinally for the budget timelines and cross-sectionally to enable comparisons between different countries.

Both purposes have shaped the development of the QPF concept. As an example, the author of this paper conducted in 2002–03 a first study on the quality of public finances of the German federal budget and the aggregate public finances of Germany on behalf of the Federal Ministry of Finance. The aim of the research was to develop an indicator of future-oriented expenditure that would serve this purpose better than the benchmark traditionally used, that is public investment expenditure. The result was the ‘WNA budget’ (Thöne 2005), with which the growth- and sustainability-effective expenditures (*WNA: Wachstums- und Nachhaltigkeitswirksame Ausgaben*) were shown separately from the other government expenditures.

Based on an intergenerational understanding of sustainability, spending in the following areas was included in the ‘WNA Budget’: Kindergartens, schools, universities and other education, science and R&D outside universities, measures of family policy, active labour market policy, health, environment and nature conservation as well as the promotion of renewable energies. In addition, expenditure on infrastructure services in transport and communications was included. The study also showed that a delineation of QPF expenditure solely on the basis of the Functional Statistics for Governmental Activities (COFOG) offers the simplest implementation.

Apart from being used in the domestic discussion of future fiscal governance, the *WNA Budget* served as an early input to the *Working Group on the Quality of Public Finances* WGQPF of the Economic Policy Committee of the European Union which, subsequently, has developed the whole spectrum of the quality discussion (see above).

With regard to the composition of public expenditure, this group has achieved two key results. For one, it has provided the common ground for the EU Commission to present a proposal for a QPF indicator that was to be applicable to all member states (EU-COM 2008). Once more, the practical feasibility of this approach was demonstrated by using the German example (Thöne and Dobroschke 2010). However, this EU initiative proved to be far too ambitious and politically too invasive to gain an important role in the EU’s fiscal governance. To be effective, a QPF approach for all member states would have had to be anchored in the preventive arm of the Stability and Growth Pact. But this would have created more European influence on the budget priorities of the member states than they would have wanted to concede. As a consequence, only rather soft quality criteria could be implemented within the corrective arm of the Excessive Deficit Procedure.

Politically, the quality of public finances has so far not been able to develop the authority at the European level that it would deserve, given its high significance for Europe’s future. This makes the second achievement of the WGQPF on the composition of public finances all the more important and enduring. Already the first studies had shown that the functional analysis of government expenditure was the most practical way to implement the quality concept. However, no data were available in sufficient detail. Specifically, for the EU states only insufficiently rough expenditure data according to the international COFOG classification were available. As a result of the initiative of the WFQPF, from 2005 onwards a task force was set up at Eurostat which, with and for the EU Member States, compiled data on public expenditure at the reasonably detailed level of the COFOG 2-digit-classifier, retroactively from *de facto* 1995 onwards (Eurostat 2011). Without the WGQPF impetus, these data, which are indispensable for the QPF-discussion in Europe, would most likely not be collected even today.

Nevertheless, it has to be acknowledged in retrospect that the world economic crisis from autumn 2008 and the resulting European sovereign debt crisis of the years 2010–12 washed away the discussion on modernisation of the quality of public finances in all its aspects.

All of a sudden, it was a matter of averting the collapse of the financial system, counteracting the ensuing recession in the real economy, saving some EU states from bankruptcy and bringing the common currency, the euro, through this fire baptism reasonably unscathed. During the crisis, little attention was paid to the discussion on the quality of the longer-term future orientation and rationality of fiscal policy. The focus of fiscal policy was once again on simple quantitative issues. Qualitative aspects, questions about the investive or sustainable character of government spending, played no role for the time being.

Until the next major crisis, which erupted in 2020 with the COVID-19 pandemic, the concept of quality of public finances has not experienced significant political progress at the level of the European Union or in individual countries in or outside the European Union. However, academic attention has not faltered, so the next steps for advancing the concept are quite clear.

In the course of ten years it has become apparent that the primary goals of a quality philosophy cannot remain unchanged over time. This became clear in the above discussion with regard to the relationship between economic growth and the goal of climate protection and resource conservation. A modern, future-oriented QPF concept must give top priority to the finite nature of natural resources and the correspondingly decoupling-focused concept of economic growth. Crisis resilience and the question of how inclusive growth is (Cournede, Fournier and Hoeller 2018) will also play a much greater role in the 2020s than in the previous decade.

Also, the 'framing' of the QPF discussion has changed with the COVID-19 crisis. The long phase of extremely low interest rates has reduced concerns about fiscal sustainability in many countries – despite the enormous new debt in the wake of COVID-19. However, it would be a mistake to conflate the idea of quality with its merely temporal genesis with a policy of strict fiscal austerity, which seems to have been overcome at the beginning of the 2020s. In an era that seeks to make future-oriented policies possible again through future-oriented financing – that is through public debt – the importance of the QPF concept does not diminish, but rather increases.

Just like national fiscal rules, the European Stability and Growth Pact is also under scrutiny. Here, the focus is no longer only on the quality of public spending, but increasingly also on the quality of new public debt. Modern fiscal rules can no longer be simple numerical deficit limits. Rather, they specify the permissible level of new debt according to the purpose of the expenditure financed by it. This does not mean that all spending covered by the QPF concept should be financed by deficits. Teachers' salaries must always be financed by current taxes. Nevertheless, analogous to the traditional 'golden rule' of public debt, some government expenditures with particularly high benefits for the future can also be financed by the future through debt. Especially in the case of climate protection, such a 'swap of future burdens' – low additional burden from debt, substantial saved burden on the part of the climate – can become an element of a concept that is more strongly based on the quality of public finances, for example a 'green golden rule'.

However, this QPF perspective on the funding of public tasks should also be reinforced for other elements of public finances that are particularly effective in terms of 'future investment'. In the best case, such a perspective is embedded in a redesign of fiscal governance that relies more on soundly interpreted economic standards than on rigid numerical rules.[1] Such a comprehensive, more economic governance would also benefit immensely from a strong reliance on the broad knowledge on public sector efficiency as discussed in the various chapters of this handbook.

NOTE

1. See for example Wieser (2021) and Blanchard/Leandro/Zettelmeyer (2021).

REFERENCES

Afonso, A., W. Ebert, L. Schuknecht and M. Thöne (2005), *Quality of Public Finances and Growth*, ECB Working Paper No. 438, European Central Bank, Frankfurt/M.
Afonso, A. and J. G. Alegre (2011), Economic growth and budgetary components: a panel assessment for the EU. *Empirical Economics*, 41: 703–723.
Agénor, P. and O. Canuto (2012), *Measuring the Effect of Gender-Based Policies on Economic Growth,* The World Bank. Economic Premise No. 85, June 2012.
Agénor, P. and O. Canuto (2014), Gender equality and economic growth in Brazil: A long-run analysis. *Journal of Macroeconomics* 43, Issue C, 155–172.
Alders, P. and P. D. Broer (2004), Ageing, fertility and growth. *Journal of Public Economics*, 89: 1075–1095.
Barro, R. (2013), Education and economic growth. *Annals of Economics and Finance*, 14 (2): 329–366.
Barro, R. and X. Sala-i-Martin (1995), *Economic Growth*. McGraw-Hill; New York.
Bertenrath, R., M. Thöne and C. Walther (2006), *Wachstumswirksamkeit von Verkehrsinvestitionen in Deutschland*, FiFo-Berichte Nr. 7, Köln.
Blanchard, O., Á. Leandro and J. Zettelmeyer (2021), *Redesigning EU Fiscal Rules: From Rules to Standards*, PIIE Working Paper 21–1; Washington D.C.
Bloom, D. E., D. Canning and J. Sevilla (2001), *The Effect of Health on Economic Growth: Theory and Evidence*, NBER Working Paper No. 8587, Cambridge MA.
Bloom, D., D. Canning, R. Kotschy, K. Prettner and J. Schünemann (2019), *Health and Economic Growth: Reconciling the Micro and Macro Evidence*, NBER Working Paper No 26003, Cambridge MA.
Cepparulo, A. and G. Mourre (2020), *How and How Much? The Growth-Friendliness of Public Spending through the Lens*, European Economy Discussion Paper 132, Brussels.
Colombier, C. (2011), *Does the composition of public expenditure affect economic growth?, Evidence from the Swiss case*, Applied Economics Letters 18 (16): 1583–1589.
Cournede, B., J.-M. Fournier, P. Hoeller (2018), *Public finance structure and inclusive growth*, OECD Economic Policy Paper No. 25, Paris.
Delalibera, B. R. and P. C. Ferreira (2019), Early childhood education and economic growth. *Journal of Economic Dynamics and Control* 98: 82–104.
Deroose, S. and C. Kastrop (Hrsg.) (2008), *The Quality Of Public Finances*, Occasional Papers 37, 2008, European Commission, Directorate-General for Economic and Financial Affairs. Brussels.
Diamond, A. R. (1999), Does federal funding 'Crowd In' private funding of science? *Contemporary Economic Policy*, 17: 423–431.
Dollar, D. and Gatti, R. (1999), *Gender Inequality, Income and Growth: Are Good Times Good for Women?* Policy Research Department on Gender and Development. Working Paper No. 1.
Esfahani, H. and M. T. Ramíres (2003), Institutions, infrastructure and economic growth. *Journal of Development Economics,* 70: 443–477.
European Commission (EU-COM) (2003), Public Finances in EMU 2003. Brussels.
European Commission (EU-COM) (2008), Public Finances in EMU 2008. Brussels.
European Commission (EU-COM) (2014), *Infrastructure in the EU: Developments and Impact on Growth*, Occasional Papers 203. Brussels.
Eurostat (2011), *Manual on sources and methods for the compilation of COFOG statistics - Classification of the Functions of Government (COFOG)* – 2nd edition, Luxembourg.
Fedderke, J. and R. Garlick (2008), *Infrastructure Development and Economic Growth in South Africa: A Review of the Accumulated Evidence*, Policy Paper No 12. School of Economics, University of Cape Town and Economic Research Southern Africa.
Fritschi, T. and T. Oesch (2008), *Volkswirtschaftlicher Nutzen von frühkindlicher Bildung in Deutschland. Eine ökonomische Bewertung langfristiger Bildungseffekte des Besuchs von Kindertageseinrichtungen*, Studie im Auftrag der Bertelsmann Stiftung, Bern.
Guellec, D. and B. van Pottelsberghe de la Potterie (2000), *The Impact of Public R&D Expenditure on Business R&D*. OECD STI Working Paper 4.
Heckman, J. and F. Cunha (2007), The technology of skill formation. *American Economic Review* 97 (2), 31–47.
Heinemann, F., A. Bohne, P. Breithaupt, T. Doherr, G. Licht, T. Niebel, M. Thöne (2021), *Studie zur Einführung einer Zukunftsquote.* Im Auftrag des Bundesministeriums für Bildung und Forschung, ZEW Mannheim.
IMF (2013), *Women, Work and the Economy: Macroeconomic Gains from Gender Equity*. IMF Staff Discussion Note SDN 10/13, Washington D.C.
IPCC Intergovernmental Panel on Climate Change (2021), *Sixth Assessment Report, Climate Change 2021: The Physical Science Basis*, Geneva.
Kabeer, N. and L. Natali (2013), *Gender Equality and Economic Growth: Is there a Win-Win?* IDS Working Paper 417. Institute of Development Studies, Brighton.

Kalaitzidakis, P. and S. Kalyvitis (2005), "New" Public Investment and/or Public Capital Maintenance for Growth? The Canadian Experience, *Economic Inquiry*, 43(3): 586–600.
Lorgelly, P. K. (2000), *Are there Gender-Separate Human Capital Effects on Growth? A Review of the Recent Empirical Literature*, University of Nottingham.
Lucas, R. E. (1988), On the mechanics of economic development. *Journal of Monetary Economics*, 22, 3–42.
Mincer, J. (1974), *Schooling, Experience and Earnings*. Human Behavior & Social Institutions No. 2. NBER Books, New York.
OECD (2021), *Climate Change and Long-term Fiscal Sustainability, Scoping Paper*. Working Party of Senior Budget Officials, GOV/PGC/SBO(2021)4, Paris.
Öztürk, F. and A. Altun Ada (2013), *Is Health Expenditure Important for Economic Growth in Selected EU Countries*. The Empirical Economics Letters, January.
Phang, S (2003), Strategic development of airport and rail infrastructure: the case of Singapore. *Transport Policy*, 10: 27–33.
Pradhan, R. P. and T. P. Bagchi (2013), Effect of transportation infrastructure on economic growth in India: The VECM approach. *Research in Transportation Economics*, 38: 139–148.
Prognos (2014), *Gesamtevaluation der ehe- und familienbezogenen Maßnahmen und Leistungen in Deutschland.* Auftraggeber: Bundesministerium der Finanzen und Bundesministerium für Familie, Senioren, Frauen und Jugend, Berlin.
Quirino, E., P. Macas Nunes and A. F. De Matos (2014), Public investment and economic growth in Portuguese sub-regions: Empirical evidence using panel data. *Transformations in Business and Economics*, 13 (1): 60–80.
Rodriguez-Vives, M. (2019), The quality of public finances: where do we stand? *Economics and Business Letters*, 8 (2), 97–105.
Romer, P. M. (1990), Endogenous technological change. *Journal of Political Economy*, 98(5): 71−102.
Short, J. and A. Kopp (2005), Transport infrastructure: Investment and planning policy and research aspects. *Transport Policy*, 12: 360–367.
Strauss, J. and D. Thomas (1998), Health, nutrition and economic development. *Journal of Economic Literature*, 36 (June): 766−817.
Strittmatter, A. and U. Sunde (2013), *Health and Economic Development: Evidence from the Introduction of Public Health Care*, IZA DP No. 5901.
Temple, J. (2000), *Growth Effects of Education and Social Capital in the OECD Countries*. OECD Economics Department Working Paper 263.
Thöne, M. (2005), *Wachstums- und nachhaltigkeitswirksame öffentliche Ausgaben (WNA),* FiFo-Berichte Nr. 2, Köln.
Thöne, M. and F. Krehl (2016), *Investing in Our Future. Empirical Evidence of the Impact of Public Spending on Inclusive Growth*, Bertelsmann-Stiftung, Gütersloh.
Thöne, M. and S. Dobroschke (2010), *Qualität der öffentlichen Finanzen - Anwendung des Ansatzes der EU-Kommission auf Deutschland*, FiFo-Berichte Nr. 12, Köln.
Uzawa, H. (1965), Optimal technical change in an aggregative model of economic growth. *International Economic Review*, 6 (1): 18−31.
Wang, E. C. (2002), Public infrastructure and economic growth: A new approach applied to East Asian Economies. *Journal of Policy Modelling*, 24: 411–435.
Weil, N. D. (2006), *Population Aging*. NBER Working Paper No. 12147.
Wieser, T. (2021), *The Post-Coronavirus Fiscal Policy Questions Europe Must Answer*, BRUEGEL Brussels.
Wößmann, L. and M. Piopiunik (2009), *Was unzureichende Bildung kostet – Eine Berechnung der Folgekosten durch entgangenes Wirtschaftswachstum*. Bertelsmann Stiftung, Gütersloh 2009.
Zouhar, Y., J. Jellema, N. Lustig and M. Trabelsi (2021), *Public Expenditure and Inclusive Growth - A Survey*, IMF Working Paper WP/2183, Washington D.C.

3. Government spending efficiency, measurement and applications: A cross-country efficiency dataset

António Afonso, João Tovar Jalles, and Ana Venâncio

1. INTRODUCTION

A country´s performance is, in part, dictated by the size of its public sector and the efficiency level with which it uses its (typically scarce) resources.[1] It is, therefore, important from both economic and policy points of view to evaluate the performance of the public sector and understand the determinants of public sector efficiency so as to maximize welfare but also to optimize investment projects and, in that way, propel growth forward. There has been an ongoing debate in the literature over the role and size of the government (Afonso and Schuknecht, 2019), mostly motivated by the substantial heterogeneity across countries in terms of the government spending.[2] This issue is even more relevant when governments face strict government budget constraints and most western economies have been living in the low growth phase for several years now, notably in the context of economic downturns and of scarce public resources.

In this chapter, we undertake a systematic review of the literature dealing with the overall public sector performance and efficiency, we define a methodology to compute public sector efficiency and we create a novel and large cross-country panel dataset of government indicators and public sector efficiency scores. We cover a sample of 36 Organisation for Economic Co-operation and Development (OECD) countries over the 2006–17 time period. More specifically, firstly, we start by defining a set of economic and sociodemographic metrics and we construct composite performance indicators. Previous papers on this topic have typically studied a very limited number of countries over a one- or two-year time span, which is a gap we are trying to cover with this work. Secondly, we compute and report a full set of (input and output oriented) efficiency scores on the basis of the performance indicators previously calculated, relating performance outputs and input measures of government spending.

The remainder of the chapter is organized as follows. Section 2 reviews the relevant literature. Section 3 presents some of methods used to obtain public sector efficiency measures. Section 4 discusses recent empirical applications. The last section concludes.

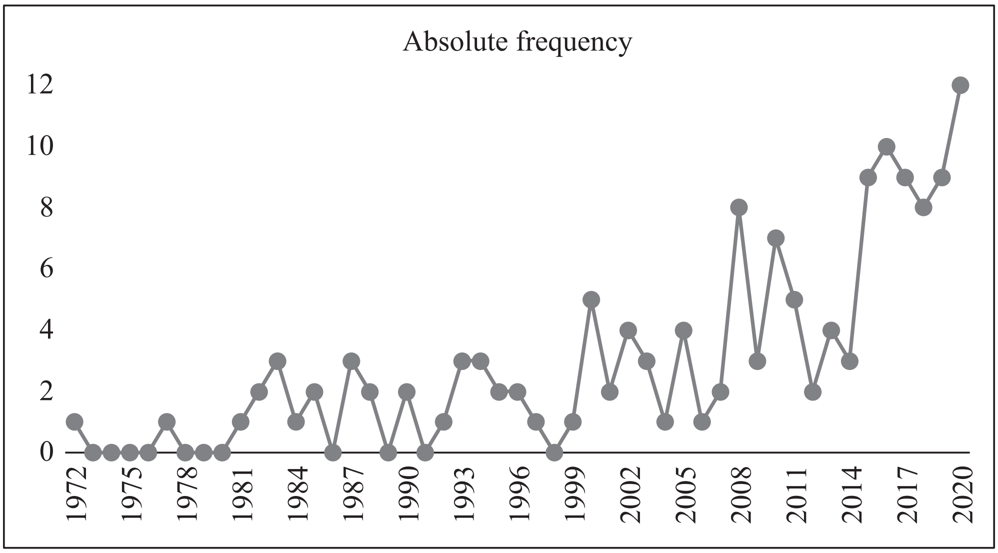
2. LITERATURE REVIEW

The efficient provision of services and goods by governments has become one of the key issues discussed in the public finance literature over the last 20 years (see for example the works by Gupta and Verhoeven, 2001; Tanzi and Schuknecht, 1997, 2000; Afonso et al., 2005).

In this section, we review the main studies on public sector efficiency by applying the following methodology. We search the Web of Science[3] for English language articles published after

1970 in academic, peer-reviewed journals. To identify relevant publications, we searched for works using two queries: i) with 'public sector efficiency' in the tile, and ii) with 'public sector' or 'efficiency' in the title and 'public sector efficiency' in the title, text, abstract or keywords. The exact search strings were: i) TI = (public sector efficiency) and ii) ALL = 'public sector efficiency' AND TI = (public sector OR efficiency). As a result of the search, a total of 142 and 55 articles were identified for queries i) and ii), respectively. Then, we screened these articles to evaluate the topic fit and eliminated those that evaluated local government performance and the performance of a specific public service provided by the local and central governments.[4] In doing this, we also evaluated the study subject, research question and findings.

Figure 3.1 shows the number of publications published per year, using both sets of queries. We observe an increasing trend in publications since 2000, with peaks in the period 2008–10 and in the period 2019–20. This reflects the growing interest of academic research in this particular area, which may have been prompted notably by the fiscal institutional set-up, for example, in the European Union. Indeed, after the creation of the Economic and Monetary Union in the European Union in the early 1990s, enhanced fiscal coordination and surveillance ensued, with increased awareness of the relevance of fiscal sound behaviour. In addition, the driver and the need to implement fiscal consolidations in the European Union (due to the convergence criteria that needed to be met) raised the bar in terms of assessing how much and what quality of public services the government are providing, while economic crisis also shed attention of the use of scarce public resources. Hence, both performance and efficiency started playing a bigger role in the 2000s in the EU case.[5]



Source: Web of Science and own elaboration.

Figure 3.1 Yearly publications on the topic of public sector efficiency in Web of Science

3.1.a Query with public sector efficiency in the title

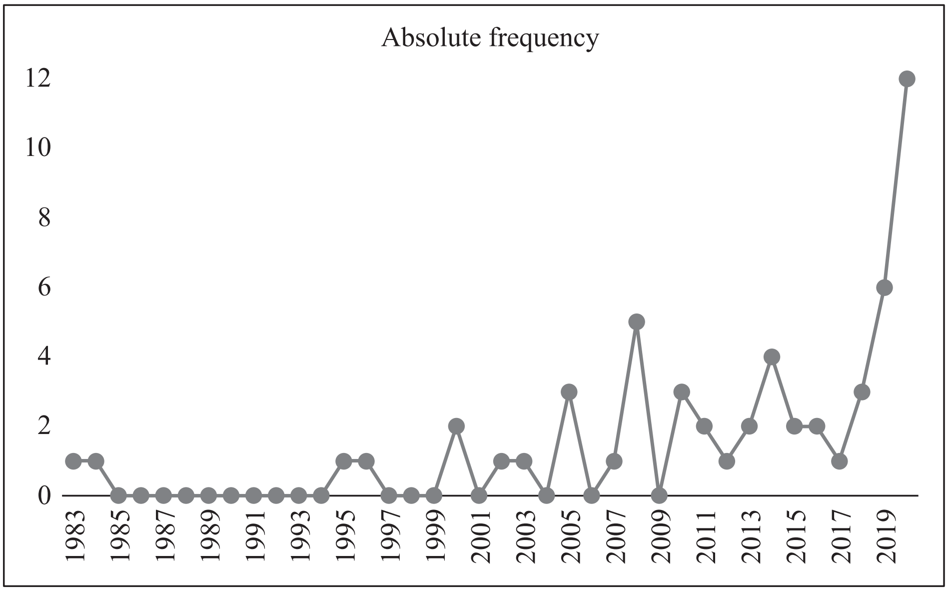


Figure 3.1 (Continued)

3.1.b Query with public sector efficiency in the title, text, abstract and keywords

Journals that more frequently show up in the abovementioned sample extractions are *Applied Economics*, *European Journal of Operational Research*, *European Journal of Political Economy*, *Journal of Public Economics* and *Public Choice*.

Several studies assess public sector efficiency looking at different sample and time spans but most tend to focus on OECD and European countries (Adam at al., 2011; Dutu and Sicari, 2016; Afonso and Kazemi, 2017; Antonelli and de Bonis, 2019). Much less evidence is available about government relative efficiency in other areas of the world such as Africa, Asia or Latin America. That said, some studies report some first empirical explorations for Latin American and Caribbean countries (see for example Afonso et al., 2013).

Two key results emerge from this literature: i) public spending efficiency can be improved; and ii) specific factors are associated with efficiency. These cross-country aggregated efficiency studies are very useful to compare the performance of different countries, nevertheless, it is important to take into account the underlying institutional, cultural, political and economic factors (Mandl et al., 2008). To account for these issues, studies have resorted to two-stage models.[6] Results suggest that education, income level, quality of the institutions and a country's governance are positively and statistically significantly associated with performance (Afonso and St. Aubyn, 2005; Hauner and Kyobe, 2008; Antonelli and de Bonis, 2019). Others report that political variables, such as having a right-wing and a strong government and also high voter participation rates and decentralization of the fiscal systems, are positively associated with more efficient public sectors (Adam et al., 2011). More recently, Afonso et al. (2019, 2020) evaluated the role of tax structures and tax reforms on explaining cross-country efficiency differences. Table 3.1 provides a short summary of results of these papers assessing overall public sector performance and efficiency.

Table 3.1 Overall public sector efficiency

Authors	Sample	Methods	Results
Afonso et al. (2005)	23 OECD countries	FDH	The average input efficiency score of the 15 EU countries is 0.73 (around 27% could be reduced).
Adam et al. (2011)	19 OECD countries, 1980−2000	Stochastic DEA	Countries with right-wing and strong governments, high voter participation rates and decentralized fiscal systems, are expected to have higher PSE.
Afonso et al. (2013)	Latin American and Caribbean countries, 2001−10	DEA	Output efficiency scores higher than input efficiency scores. PSE is inversely correlated with the size of the government, while the efficiency frontier is defined by Chile, Guatemala and Peru.
Dutu, Sicari (2016)	35 OECD countries, 2012	DEA	Wide dispersion in efficiency measures across OECD, health care, education and general administration.
Chan et al. (2017)	115 countries	Panel GMM	VAT system enhances the effect of efficient government spending on the economic growth.
Herrera, Ouedrago (2018)	175 countries for 2006–16 on education, health, infrastructure	FDH, DEA	The efficiency of capital spending is correlated with regulatory quality and perception of corruption.
Mohanty, Bhanumurthy (2018)	27 Indian States, 2000−15	DEA	Higher efficiency on education than on health and overall social spending. Governance and growth affects the efficiency.
Montes et al. (2019)	68 developing and 14 developed countries, 2006–14	Panel, GMM	Fiscal transparency affects government spending efficiency.
Antonelli, de Bonis (2019)	22 EU countries, 2013	Median voter model	More efficient have higher education and GDP levels, smaller population size, lower degree of selectivity of their welfare systems and a lower corruption level.

Notes: DEA, data envelopment analysis; FDH, free disposal hull; GMM, Generalized Method of Moments.

3. DATA AND VARIABLES

Our novel dataset includes 36 OECD countries[7] for the period between 2006 and 2017. We gather data from several publicly available sources, such as the World Economic Forum, World Bank, World Health Organization, IMF World Economic Outlook and OECD database. When data was not available for a specific year, we assumed that the data was equal to that of the previous year.

Government spending can have many (often competing) objectives (promoting stability, allocation and redistribution) and any definition of efficiency must be understood in this Musgravian sense. Following the related literature, we use a set of metrics to construct a composite indicator of Public Sector Performance (PSP), as suggested by Afonso et al. (2005, 2019). PSP is then computed as the average between opportunity and Musgravian indicators. In that vein, we distinguish the role of government in providing 'opportunities' and a level playing field in the markets and the traditional Musgravian functions of the government.

First, opportunity indicators reflect governments' performance in the administration, education, health and infrastructure sectors. The administration sub-indicator includes the following measures: corruption, burden of government regulation (red tape), judiciary independence, shadow economy and property rights. To measure the education sub-indicator, we use the secondary school enrolment rate, quality of educational system and Programme for International Student Assessment (PISA) scores. For the health sub-indicator, we compile data on the infant survival rate, life expectancy and survival rate from cardiovascular diseases (CVD), cancer, diabetes or chronic respiratory diseases (CRD). The infrastructure sub-indicator is measured by the quality of overall infrastructure.

Second, Musgravian indicators assess governments' performance in allocation, distribution and stabilization. Our socio-economic proxies for performance include three sub-indicators: distribution, stability and economic performance. To measure income distribution and inequality, we use the Gini coefficient. For the stability sub-indicator, we use the coefficient of variation for the 5-year average of Gross Domestic Product (GDP) growth and the rolling overlapping standard deviation of 5 years inflation rate. To measure economic performance, we include the 5-year average of real GDP per capita, real GDP growth and unemployment rate. Accordingly, both opportunity and Musgravian indicators result from the average of the measures included in each sub-indicator. To ensure a convenient benchmark, each sub-indicator measure is normalized by dividing the value of a specific country by the average of that measure for all countries in the sample. Table 3.2 lists all sub-indicators to construct the PSP indicators and provides further information on the sources and variable construction.

Table 3.2 DEA output components

Sub Index	Variable	Source	Series
Opportunity Indicators			
Administration	Corruption	Transparency International's Corruption Perceptions Index (CPI) (2006−17)	Corruption on a scale from 10 (Perceived to have low levels of corruption) to 0 (highly corrupt), 2006–11; Corruption on a scale from 100 (Perceived to have low levels of corruption) to 0 (highly corrupt), 2012–17.
	Red Tape	World Economic Forum: The Global competitiveness Report (2006–17)	Burden of government regulation on a scale from 7 (not burdensome at all) to 1 (extremely burdensome).

Table 3.2 (Continued)

Sub Index	Variable	Source	Series
	Judicial Independence	World Economic Forum: The Global competitiveness Report (2006–17)	Judicial independence on a scale from 7 (entirely independent) to 1 (heavily influenced).
	Property Rights	World Economic Forum: The Global competitiveness Report (2006–17)	Property rights on a scale from 7 (very strong) to 1 (very weak).
	Shadow Economy	Schneider (2016) (2006–16)[8]	Shadow economy measured as percentage of official GDP. Reciprocal value $1/x$.
Education	Secondary School Enrolment	World Bank, World Development Indicators (2006–17)	Ratio of total enrolment in secondary education.
	Quality of Educational System	World Economic Forum: The Global competitiveness Report (2006–17)	Quality of educational system on a scale from 7 (very well) to 1 (not well at all).
	PISA Scores	PISA Report (2003, 2006, 2009, 2012, 2015)	Simple average of mathematics, reading and science scores for the years 2015, 2012, 2009; Simple average of mathematics and reading for the year 2003. For the missing years, we assumed that the scores were the same as in the previous years.
Health	Infant Survival Rate	World Bank, World Development Indicators (2006–17)	Infant survival rate = (1000−IMR)/1000. IMR is the infant mortality rate measured per 1000 lives birth in a given year.
	Life Expectancy	World Bank, World Development Indicators (2006–17)	Life expectancy at birth, measured in years.
	CVD, Cancer, Diabetes or CRD Survival Rate	World Health Organization, Global Health Observatory Data Repository (2000, 2005, 2010, 2015, 2016)	CVD, cancer and diabetes survival rate = 100 − M. M is the mortality rate between the ages 30 and 70. For the missing years, we assumed that the scores were the same as in the previous years.
Public Infrastructure	Infrastructure Quality	World Economic Forum: The Global competitiveness Report (2006–17)	Infrastructure quality on a scale from 7 (extensive and efficient) to 1 (extremely underdeveloped).

(Continued)

Table 3.2 (Continued)

Sub Index	Variable	Source	Series
Standard Musgravian Indicators			
Distribution	Gini Index	Eurostat, OECD (2006–16)[9]	Gini index on a scale from 1 (perfect inequality) to 0 (perfect equality). Transformed to 1-Gini.
Stabilization	Coefficient of Variation of Growth	IMF World Economic Outlook (WEO database) (2006–17)	Coefficient of variation=standard deviation/mean of GDP growth based on 5-year data. GDP constant prices (% change). Reciprocal; value 1/*x*.
	Standard Deviation of Inflation	IMF World Economic Outlook (WEO database) (2006–17)	Standard deviation of inflation based on 5-year consumer prices (% change) data. Reciprocal value 1/*x*.
Economic Performance	GDP per Capita	IMF World Economic Outlook (WEO database) (2006–17)	GDP per capita based on Purchasing Power Parity (PPP), current international dollar.
	GDP Growth	IMF World Economic Outlook (WEO database) (2006–17)	GDP constant prices (% change).
	Unemployment	IMF World Economic Outlook (WEO database) (2006–17)	Unemployment rate, as a percentage of total labour force. Reciprocal value 1/*x*.

Table 3.3 Input components

Sub Index	Variable	Source	Series
Opportunity Indicators			
Administration	Government Consumption	IMF World Economic Outlook (WEO database) (2005–16)	General government final consumption expenditure (% of GDP) at current prices.
Education	Education Expenditure	UNESCO Institute for Statistics (2005–16)[10]	Expenditure on education (% of GDP).
Health	Health Expenditure	OECD database (2005–16)	Expenditure on health (% of GDP).
Public Infrastructure	Public Investment	European Commission, AMECO (2005–16)[11]	General government gross fixed capital formation (% of GDP) at current prices.
Standard Musgravian Indicators			
Distribution	Social Protection Expenditure	OECD database (2005–16)[12]	Aggregation of the social transfers (% of GDP).
Stabilization/ Economic Performance	Government Total Expenditure	OECD database (2005–16)[13]	Expenditure total expenditure (% of GDP).

Note: AMECO is the annual macro-economic database of the European Commission's Directorate General for Economic and Financial Affairs.

Our input measure, Public Expenditure (PE), is expressed as a percentage of GDP and it considers each area of government expenditure. More specifically, we consider government consumption as input for administrative performance, government expenditure in education as input for education performance, health expenditure as input for health performance and public investment as input for infrastructure performance. For the distribution indicator, we consider expenditures on transfers and subsidies. The stability and economic performance are related to the total expenditure. Table 3.3 includes data on various governments' expenditures and provides further information on the sources and variable construction.

Tables 3.4 and 3.5 show the evolution of the standardized PSP and PE indicators, respectively, normalized to one in each year. For instance, the overall dispersion of the PSP indicator, although not too different between 2006 and 2017, increased during the European debt crisis of 2011–13. Note that Greece presented a negative performance on the stability and economic performance sub-indicators in years 2012 and 2013 and, consequently, the Musgravian and the overall PSP score are negatives. In addition, in 2012 and 2013 we witness a higher volatility in the PSP indicator for the full country sample, which can be linked to the changes in major economic indicators in the aftermath of the Global and Financial Crisis.

Table 3.4 PSP standardized indicator

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AUS	1.11	1.09	1.27	1.35	1.45	1.53	2.16	1.94	1.22	1.19	1.17	1.28
AUT	1.12	1.09	1.21	1.09	1.09	1.10	1.07	0.97	1.03	1.00	1.02	1.01
BEL	0.99	1.00	1.01	0.98	1.02	1.02	0.98	0.96	0.96	0.96	0.98	0.97
CAN	1.13	1.08	1.31	1.05	1.10	1.10	1.19	1.17	1.24	1.08	1.13	1.09
CHE	1.21	1.23	1.35	1.20	1.25	1.25	1.28	1.34	1.33	1.29	1.27	1.19
CHL	0.87	0.87	1.10	0.89	1.03	1.07	1.43	1.30	0.94	0.92	0.85	0.81
CZE	0.91	0.94	1.09	0.94	0.92	0.90	0.76	0.77	0.91	0.96	0.91	0.97
DEU	1.00	0.98	1.09	1.03	1.08	1.08	1.05	1.02	1.09	1.04	1.06	1.06
DNK	1.09	1.07	0.99	1.08	1.04	1.00	0.96	0.96	1.06	1.05	1.05	1.16
ESP	1.19	1.15	1.03	0.89	0.83	0.76	0.41	0.37	0.78	0.83	0.85	0.85
EST	1.00	0.99	0.40	0.95	0.79	0.91	1.02	0.86	0.94	0.89	0.93	0.94
FIN	1.05	1.07	1.05	1.07	1.05	1.04	0.84	0.89	0.95	0.91	0.97	0.99
FRA	1.14	1.06	1.01	0.99	1.00	1.01	0.96	0.97	1.00	0.97	0.99	1.01
GBR	1.11	1.08	0.89	1.02	1.00	0.96	1.02	1.02	1.11	1.07	1.07	1.02
GRC	0.95	0.95	0.82	0.87	0.61	0.42	−0.22	−0.12	0.62	0.64	0.67	0.71
HUN	1.01	0.76	0.82	0.83	0.75	0.75	0.60	0.76	0.86	0.83	0.80	0.91
IRL	1.02	1.02	0.65	0.91	0.87	0.91	0.80	0.93	1.11	1.43	1.06	1.19
ISL	1.05	1.12	1.14	1.05	0.84	0.94	0.91	1.01	0.99	1.11	1.18	1.09
ISR	0.84	0.89	1.09	1.01	1.21	1.28	1.49	1.55	1.03	0.98	1.07	1.06
ITA	0.92	0.89	0.73	0.84	0.82	0.77	0.44	0.55	0.73	0.73	0.73	0.80
JPN	1.03	1.20	0.89	0.99	1.04	0.94	1.04	1.10	0.99	1.03	1.04	1.01
KOR	1.00	1.05	1.23	1.15	1.32	1.29	1.49	1.43	1.06	1.18	1.29	1.37

(Continued)

Table 3.4 (Continued)

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
LTU	0.92	0.95	0.97	0.90	0.74	0.85	0.96	0.87	0.92	0.89	0.94	0.90
LUX	1.15	1.16	0.98	1.12	1.19	1.12	0.99	1.21	1.19	1.08	1.09	1.08
LVA	0.95	0.96	0.44	0.87	0.57	0.78	0.87	0.76	0.79	0.88	0.92	0.87
MEX	0.80	0.79	0.87	0.85	0.93	0.90	1.14	0.93	1.08	1.00	0.94	0.81
NLD	0.98	1.09	1.35	1.23	1.17	1.13	0.94	0.90	0.99	1.01	1.00	1.03
NOR	1.04	1.06	1.07	1.13	1.09	1.09	1.30	1.19	1.16	1.17	1.07	1.09
NZL	1.03	1.03	0.92	1.09	1.14	1.13	1.35	1.55	1.20	1.11	1.24	1.09
POL	0.80	0.82	1.12	1.04	1.21	1.38	1.63	1.31	0.90	0.89	0.90	0.91
PRT	0.83	0.86	0.97	0.86	0.87	0.74	0.35	0.53	0.78	0.80	0.82	0.91
SVK	0.90	0.89	1.23	0.92	0.96	0.89	0.95	0.84	0.87	0.91	0.90	0.90
SVN	0.94	0.94	1.19	0.99	0.89	0.84	0.54	0.67	0.88	0.84	0.86	0.91
SWE	1.04	1.05	0.92	1.05	1.18	1.10	0.96	1.07	1.05	1.17	1.14	1.03
TUR	0.80	0.76	0.79	0.81	0.98	1.06	1.22	1.31	0.99	0.97	0.93	0.89
USA	1.06	1.06	1.01	0.96	0.98	0.95	1.11	1.10	1.25	1.18	1.15	1.10
Average	1	1	1	1	1	1	1	1	1	1	1	1
Median	1.01	1.02	1.01	0.99	1.01	1.01	0.98	0.97	0.99	0.99	0.99	1.01
Min	0.80	0.76	0.40	0.81	0.57	0.42	−0.22	−0.12	0.62	0.64	0.67	0.71
Max	1.21	1.23	1.35	1.35	1.45	1.53	2.16	1.94	1.33	1.43	1.29	1.37
Standard deviation	0.11	0.12	0.22	0.12	0.19	0.21	0.41	0.36	0.16	0.16	0.14	0.14

Source: Authors' calculations.

Table 3.5 PE standardized indicator

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AUS	0.84	0.84	0.84	0.86	0.80	0.80	0.81	0.81	0.82	0.84	0.85	0.86
AUT	1.22	1.21	1.21	1.19	1.17	1.17	1.16	1.16	1.18	1.19	1.18	1.20
BEL	1.15	1.12	1.13	1.14	1.13	1.12	1.16	1.19	1.20	1.20	1.19	1.19
CAN	0.97	0.98	1.00	0.96	0.98	1.00	0.99	0.98	0.96	0.94	0.98	1.00
CHE	0.92	0.88	0.85	0.83	0.81	0.82	0.84	0.86	0.87	0.87	0.88	0.91
CHL	0.54	0.53	0.55	0.57	0.56	0.55	0.55	0.56	0.58	0.60	0.63	0.63
CZE	1.00	0.99	0.96	0.95	0.97	0.95	0.95	0.96	0.94	0.94	0.96	0.89
DEU	1.14	1.12	1.09	1.06	1.09	1.08	1.06	1.06	1.06	1.06	1.06	1.09
DNK	1.34	1.32	1.33	1.29	1.32	1.35	1.37	1.37	1.37	1.35	1.34	1.35
ESP	0.95	0.96	0.99	1.00	1.03	1.04	1.04	1.02	0.99	0.98	0.98	0.95
EST	0.85	0.86	0.89	0.99	1.05	0.94	0.90	0.96	0.93	0.92	0.96	0.98
FIN	1.22	1.20	1.18	1.17	1.22	1.22	1.25	1.32	1.35	1.37	1.36	1.38

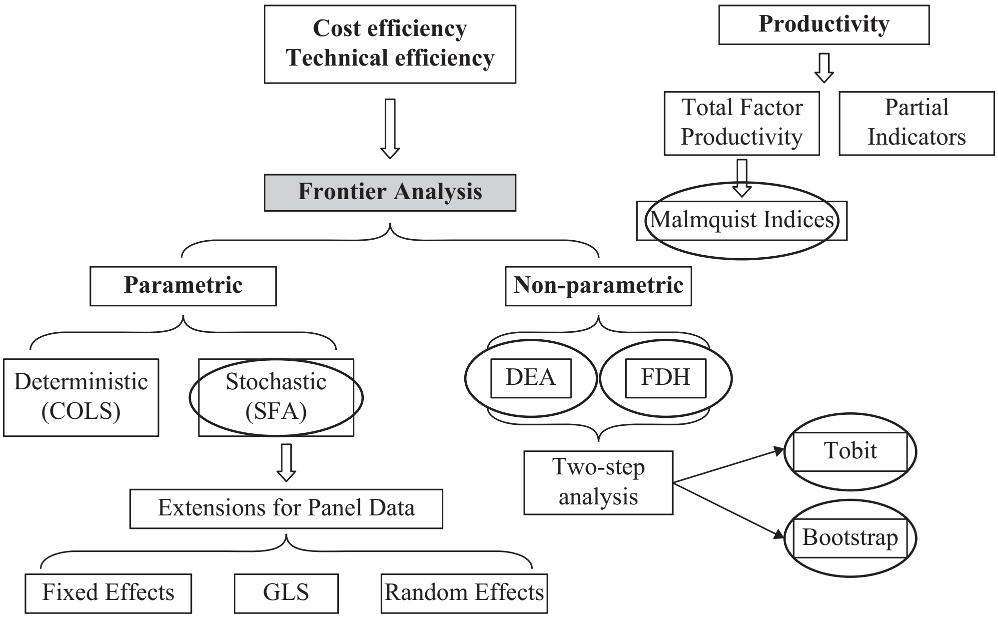
Table 3.5 (Continued)

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
FRA	1.32	1.32	1.33	1.30	1.28	1.29	1.30	1.32	1.33	1.33	1.31	1.35
GBR	0.93	0.97	0.98	1.01	1.01	1.03	1.03	1.03	1.02	1.01	1.00	1.01
GRC	1.06	1.10	1.10	1.15	1.14	1.08	1.10	1.12	1.15	1.09	1.12	1.11
HUN	1.15	1.21	1.16	1.08	1.02	1.02	1.02	1.02	1.04	1.07	1.11	0.98
IRL	0.88	0.90	0.98	1.10	1.08	1.23	1.04	0.98	0.93	0.87	0.70	0.70
ISL	1.05	1.07	1.08	1.09	0.97	0.95	0.93	0.95	0.95	0.97	0.93	1.04
ISR	0.97	0.97	0.95	0.92	0.85	0.85	0.85	0.87	0.87	0.86	0.86	0.86
ITA	1.10	1.11	1.10	1.10	1.09	1.07	1.07	1.09	1.10	1.10	1.09	1.09
JPN	0.92	0.91	0.92	0.91	0.93	0.94	1.00	1.01	1.02	1.01	0.99	1.01
KOR	0.54	0.57	0.58	0.60	0.59	0.57	0.59	0.60	0.61	0.63	0.64	0.65
LTU	0.81	0.84	0.91	0.94	1.00	0.95	0.94	0.84	0.81	0.80	0.80	0.80
LUX	1.15	1.04	1.01	1.03	1.06	1.06	1.02	1.03	1.00	0.99	0.99	1.02
LVA	0.79	0.85	0.85	0.87	0.93	0.94	0.90	0.90	0.89	0.86	0.88	0.86
MEX	0.42	0.42	0.44	0.45	0.43	0.43	0.44	0.44	0.43	0.44	0.44	0.44
NLD	1.04	1.08	1.07	1.06	1.07	1.09	1.10	1.11	1.10	1.09	1.07	1.07
NOR	1.11	1.07	1.11	1.05	1.11	1.09	1.09	1.10	1.14	1.19	1.26	1.34
NZL	0.99	1.02	1.01	0.98	0.96	1.00	1.11	1.00	0.95	0.94	0.93	0.92
POL	1.06	1.08	1.06	1.06	0.99	1.04	1.03	0.99	0.98	1.00	0.98	0.98
PRT	1.11	1.07	1.06	1.05	1.08	1.14	1.09	1.04	1.06	1.06	1.04	0.99
SVK	0.95	0.92	0.89	0.86	0.93	0.92	0.92	0.91	0.93	0.96	1.07	0.96
SVN	1.10	1.10	1.06	1.06	1.07	1.10	1.12	1.11	1.20	1.13	1.09	1.03
SWE	1.33	1.30	1.27	1.24	1.22	1.19	1.24	1.30	1.31	1.30	1.27	1.31
TUR	0.80	0.80	0.79	0.77	0.73	0.69	0.68	0.72	0.72	0.71	0.72	0.73
USA	0.88	0.88	0.91	0.94	0.93	0.91	0.91	0.88	0.85	0.94	0.95	0.97
Average	1	1	1	1	1	1	1	1	1	1	1	1
Median	0.99	1.00	1.00	1.02	1.01	1.03	1.03	1.00	0.99	0.98	0.99	0.98
Min	0.42	0.42	0.44	0.45	0.43	0.43	0.44	0.44	0.43	0.44	0.44	0.44
Max	1.34	1.32	1.33	1.30	1.32	1.35	1.37	1.37	1.37	1.37	1.36	1.38
Stdev	0.21	0.20	0.20	0.19	0.19	0.20	0.20	0.20	0.21	0.21	0.21	0.21

Source: Authors' calculations.

4. METHODOLOGY

To compute efficiency, the previously surveyed papers use several parametric and non-parametric methodologies. Parametric approaches include corrected ordinary least squares (OLS) and stochastic frontier analysis (SFA). Among the non-parametric techniques, data



Source: Own elaboration.

Figure 3.2 Methods to assess efficiency

envelopment analysis (DEA) and free disposal hull (FDH) have been widely applied in the literature. Most of the studies estimate a non-parametric production function frontier and derive efficiency scores based on the relative distances of inefficient observations from the frontier. The use of a non-parametric approach has the advantage of considering, for instance, production functions with several outputs. On the other hand, in a non-parametric set-up, differences across countries are not statistically assessed, which can be considered as a drawback of such methodology. Figure 3.2 illustrates some of the possible methods available to assess efficiency.

Following the literature, in order to compute public sector efficiency scores, we use a DEA approach,[14] which compares each observation with an optimal outcome. DEA is a non-parametric technique that uses linear programming to compute the production frontier. For each country *i* out of 36 advanced economies, we consider the following function:

$$Y_i = f\left(X_i\right), \quad i = 1, \ldots, 36 \tag{3.1}$$

where *Y* is the composite output measure (Public Sector Performance, PSP) and *X* is the composite input measure (Public Expenditure, PE), namely government spending to GDP ratio.

In Equation (3.1), inefficiency occurs if $Y_i < f(X_i)$, implying that, for the observed input level, the actual output is smaller than the best attainable one.

In computing the efficiency scores, we assume variable returns to scale (VRS) to account for the fact that countries might not operate at their optimal scale.

We use two orientations: input and output orientation. The input orientation allows us to measure the proportional reduction in inputs while holding output constant. Using the input approach, efficient scores are computed through the following linear programming problem:

$$\begin{aligned} &\min_{\theta,\lambda} \theta \\ s.t.\quad &-y_i + Y\lambda \geq 0 \\ &\theta x_i - X\lambda \geq 0 \\ &I1'\lambda = 1 \\ &\lambda \geq 0 \end{aligned} \tag{3.2}$$

where y_i is a column vector of outputs, x_i is a column vector of inputs, θ is the input efficiency score, λ is a vector of constants, $I1'$ is a vector of ones, X is the input matrix and Y is the output matrix.

In Equation (3.2), θ is a scalar (that satisfies $0 \leq \theta \leq 1$) and measures the distance between a country and the efficiency frontier, defined as a linear combination of the best practice observations. With $\theta < 1$, the country is inside the frontier, it is inefficient, while $\theta = 1$ implies that the country is on the frontier and it is efficient.

Conversely the output orientation allows us to measure the proportion increase in outputs holding inputs constant. In this approach, the efficiency scores are computed through the following linear programming problem:

$$\begin{aligned} &\max_{\varphi,\lambda} \varphi \\ s.t.\quad &-\varphi y_i + Y\lambda \geq 0 \\ &x_i - X\lambda \geq 0 \\ &I1'\lambda = 1 \\ &\lambda \geq 0 \end{aligned} \tag{3.3}$$

In Equation (3.3), φ is a scalar (that satisfies $1 \leq \varphi \leq +\infty$), and $\varphi - 1$ is the proportional increase in outputs that could be achieved by each country with input quantities held constant. In (3.3), $1/\varphi$ defines the technical output efficiency score, varying between zero and one.

Both input and output approaches, deliver the same frontier in terms of the same set of efficient countries, but the magnitude of inefficiency per country may differ between the two approaches.

5. PUBLIC SECTOR PERFORMANCE AND EFFICIENCY SCORES

We performed the DEA considering three models: baseline model (Model 0), which includes only one input (PE as percentage of GDP) and one output (PSP); Model 1 uses one input, governments' normalized total spending (PE) and two outputs, the opportunity PSP and the Musgravian PSP scores; and Model 2 assumes two inputs, governments' normalized spending on opportunity and on Musgravian indicators and one output, total PSP scores.

The detailed input efficient scores are illustrated on Tables 3.6, 3.7 and 3.8. In this analysis, we exclude Mexico because the country is efficient by default,[15] and data heterogeneity is

Table 3.6 Input-oriented DEA VRS efficiency scores Model 0

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AUS	0.88	0.77	0.80	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.93	0.76	0.75	0.76
AUT	0.62	0.52	0.48	0.50	0.48	0.47	0.48	0.48	0.51	0.51	0.54	0.54
BEL	0.47	0.50	0.49	0.51	0.50	0.49	0.47	0.47	0.48	0.50	0.53	0.54
CAN	0.80	0.64	0.77	0.61	0.58	0.55	0.56	0.57	0.82	0.66	0.65	0.65
CHE	**1.00**	**1.00**	**1.00**	0.81	0.72	0.69	0.66	0.67	**1.00**	0.84	0.73	0.71
CHL	0.99	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
CZE	0.54	0.55	0.57	0.61	0.57	0.58	0.58	0.58	0.61	0.64	0.66	0.72
DEU	0.47	0.49	0.50	0.55	0.52	0.51	0.52	0.53	0.60	0.58	0.60	0.59
DNK	0.53	0.46	0.41	0.46	0.42	0.41	0.40	0.41	0.44	0.45	0.48	0.48
ESP	0.93	0.77	0.55	0.57	0.54	0.53	0.53	0.55	0.58	0.61	0.64	0.67
EST	0.64	0.65	0.62	0.58	0.53	0.58	0.61	0.58	0.62	0.65	0.66	0.65
FIN	0.52	0.51	0.46	0.51	0.46	0.45	0.44	0.42	0.43	0.43	0.46	0.46
FRA	0.60	0.44	0.41	0.45	0.44	0.43	0.43	0.42	0.45	0.45	0.48	0.47
GBR	0.79	0.64	0.56	0.58	0.55	0.53	0.54	0.54	0.65	0.61	0.63	0.64
GRC	0.51	0.50	0.50	0.50	0.49	0.51	0.50	0.50	0.50	0.55	0.56	0.57
HUN	0.49	0.43	0.47	0.53	0.55	0.54	0.54	0.54	0.55	0.56	0.57	0.65
IRL	0.66	0.63	0.56	0.52	0.51	0.45	0.53	0.57	0.71	**1.00**	0.91	0.92
ISL	0.61	0.65	0.52	0.54	0.58	0.58	0.59	0.59	0.62	0.64	0.69	0.62
ISR	0.55	0.55	0.58	0.64	0.68	0.67	0.68	0.75	0.69	0.70	0.74	0.75
ITA	0.49	0.48	0.50	0.52	0.51	0.51	0.52	0.51	0.52	0.54	0.57	0.58
JPN	0.64	0.90	0.60	0.64	0.60	0.59	0.55	0.55	0.58	0.60	0.64	0.63
KOR	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.97	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
LTU	0.66	0.65	0.60	0.61	0.56	0.57	0.59	0.66	0.71	0.75	0.79	0.80
LUX	0.70	0.73	0.54	0.58	0.54	0.52	0.54	0.54	0.73	0.63	0.64	0.63
LVA	0.68	0.65	0.65	0.66	0.60	0.58	0.62	0.62	0.64	0.69	0.72	0.74
NLD	0.52	0.59	0.79	0.67	0.54	0.51	0.50	0.50	0.54	0.56	0.59	0.60
NOR	0.54	0.54	0.50	0.58	0.51	0.51	0.51	0.51	0.62	0.53	0.51	0.48
NZL	0.59	0.55	0.54	0.61	0.59	0.56	0.50	0.65	0.78	0.66	0.69	0.70
POL	0.51	0.49	0.52	0.56	0.58	0.63	0.60	0.57	0.59	0.60	0.64	0.65
PRT	0.48	0.50	0.52	0.54	0.52	0.48	0.51	0.54	0.54	0.56	0.61	0.64
SVK	0.57	0.58	0.67	0.67	0.60	0.60	0.60	0.61	0.62	0.62	0.59	0.66
SVN	0.49	0.49	0.54	0.55	0.52	0.50	0.50	0.50	0.48	0.53	0.58	0.62
SWE	0.46	0.44	0.43	0.48	0.47	0.46	0.45	0.43	0.46	0.48	0.50	0.49
TUR	0.67	0.66	0.69	0.74	0.76	0.79	0.82	0.78	0.83	0.84	0.88	0.87
USA	0.72	0.67	0.60	0.62	0.60	0.60	0.61	0.63	0.93	0.66	0.68	0.66

Table 3.6 (Continued)

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Count	2	3	3	3	3	3	2	3	3	3	2	2
Average	0.64	0.62	0.60	0.61	0.59	0.58	0.58	0.59	0.65	0.64	0.65	0.66
Median	0.60	0.58	0.55	0.58	0.55	0.54	0.54	0.55	0.62	0.61	0.64	0.65
Min	0.46	0.43	0.41	0.45	0.42	0.41	0.40	0.41	0.43	0.43	0.46	0.46
Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Stdev	0.16	0.16	0.16	0.14	0.15	0.15	0.15	0.15	0.17	0.15	0.13	0.14

Source: Authors' calculations.

Table 3.7 Input-oriented DEA VRS efficiency scores Model 1

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AUS	0.94	0.81	0.85	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.96	0.77	0.75	0.76
AUT	0.72	0.63	0.58	0.60	0.60	0.59	0.59	0.61	0.62	0.64	0.66	0.64
BEL	0.55	0.59	0.56	0.61	0.61	0.61	0.58	0.58	0.60	0.62	0.64	0.64
CAN	0.81	0.66	0.77	0.65	0.63	0.61	0.61	0.63	0.83	0.69	0.68	0.68
CHE	**1.00**	**1.00**	**1.00**	0.83	0.75	0.74	0.72	0.71	**1.00**	0.86	0.75	0.73
CHL	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
CZE	0.55	0.55	0.59	0.66	0.61	0.64	0.64	0.66	0.70	0.71	0.69	0.78
DEU	0.60	0.62	0.61	0.69	0.64	0.63	0.63	0.65	0.70	0.69	0.70	0.69
DNK	0.60	0.53	0.47	0.54	0.51	0.50	0.49	0.50	0.51	0.53	0.55	0.53
ESP	0.94	0.79	0.57	0.63	0.61	0.61	0.62	0.70	0.76	0.80	0.79	0.83
EST	0.65	0.68	0.67	0.59	0.58	0.65	0.66	0.60	0.65	0.68	0.67	0.67
FIN	0.60	0.61	0.54	0.61	0.58	0.57	0.55	0.53	0.54	0.55	0.57	0.56
FRA	0.67	0.52	0.48	0.55	0.54	0.54	0.53	0.53	0.55	0.57	0.60	0.57
GBR	0.92	0.72	0.63	0.68	0.67	0.64	0.63	0.66	0.72	0.69	0.71	0.70
GRC	0.57	0.53	0.56	0.60	0.59	0.65	0.68	0.69	0.69	0.73	0.73	0.72
HUN	0.55	0.47	0.54	0.68	0.70	0.68	0.67	0.68	0.68	0.64	0.61	0.76
IRL	0.67	0.64	0.56	0.56	0.59	0.54	0.62	0.67	0.79	**1.00**	0.96	0.94
ISL	0.67	0.72	0.57	0.60	0.58	0.59	0.62	0.61	0.63	0.64	0.71	0.64
ISR	0.58	0.55	0.59	0.68	0.73	0.72	0.72	0.78	0.73	0.75	0.77	0.78
ITA	0.59	0.55	0.60	0.66	0.65	0.67	0.67	0.69	0.72	0.75	0.77	0.77
JPN	0.67	0.94	0.65	0.74	0.71	0.69	0.63	0.64	0.65	0.67	0.70	0.68
KOR	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
LTU	0.68	0.66	0.62	0.65	0.65	0.64	0.65	0.74	0.79	0.84	0.84	0.86
LUX	0.74	0.81	0.60	0.67	0.63	0.62	0.64	0.68	0.87	0.77	0.78	0.75
LVA	0.70	0.67	0.72	0.67	0.68	0.69	0.68	0.65	0.68	0.75	0.76	0.80

(Continued)

Table 3.7 (Continued)

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
NLD	0.56	0.61	0.79	0.70	0.58	0.56	0.56	0.57	0.59	0.62	0.64	0.63
NOR	0.57	0.58	0.51	0.61	0.55	0.57	0.57	0.56	0.64	0.55	0.53	0.50
NZL	0.62	0.56	0.57	0.62	0.60	0.56	0.54	0.67	0.82	0.68	0.73	0.71
POL	0.61	0.56	0.60	0.64	0.66	0.72	0.66	0.66	0.70	0.70	0.73	0.77
PRT	0.52	0.53	0.57	0.63	0.61	0.55	0.61	0.69	0.72	0.75	0.77	0.81
SVK	0.66	0.63	0.77	0.77	0.71	0.72	0.70	0.74	0.75	0.73	0.62	0.75
SVN	0.56	0.54	0.59	0.62	0.60	0.59	0.60	0.61	0.59	0.61	0.66	0.72
SWE	0.53	0.52	0.49	0.55	0.55	0.55	0.51	0.48	0.50	0.52	0.54	0.52
TUR	0.88	0.80	0.84	0.95	0.93	0.98	**1.00**	0.89	0.93	0.97	0.99	0.97
USA	0.87	0.79	0.73	0.69	0.60	0.61	0.61	0.64	**1.00**	0.82	0.81	0.79
Count	3	3	3	3	3	3	4	3	4	3	2	2
Average	0.70	0.67	0.65	0.68	0.66	0.66	0.66	0.68	0.73	0.72	0.73	0.73
Median	0.66	0.63	0.60	0.65	0.61	0.63	0.63	0.66	0.70	0.70	0.71	0.73
Min	0.52	0.47	0.47	0.54	0.51	0.50	0.49	0.48	0.50	0.52	0.53	0.50
Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Stdev	0.15	0.15	0.14	0.13	0.13	0.13	0.13	0.13	0.15	0.13	0.12	0.13

Source: Authors' calculations.

Table 3.8 Input-oriented DEA VRS efficiency scores Model 2

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AUS	0.98	0.89	0.86	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.97	0.88	0.85	0.86
AUT	0.67	0.63	0.65	0.64	0.62	0.61	0.62	0.63	0.63	0.63	0.61	0.62
BEL	0.61	0.58	0.59	0.58	0.55	0.56	0.55	0.55	0.57	0.58	0.59	0.59
CAN	0.80	0.74	0.78	0.76	0.73	0.71	0.73	0.74	0.84	0.79	0.74	0.75
CHE	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
CHL	0.99	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
CZE	0.54	0.56	0.58	0.61	0.57	0.58	0.58	0.58	0.61	0.64	0.66	0.72
DEU	0.68	0.65	0.67	0.67	0.63	0.62	0.66	0.69	0.70	0.72	0.68	0.70
DNK	0.65	0.61	0.58	0.59	0.55	0.55	0.51	0.51	0.54	0.56	0.56	0.57
ESP	**1.00**	**1.00**	0.57	0.57	0.54	0.53	0.55	0.59	0.62	0.66	0.67	0.69
EST	0.65	0.65	0.62	0.59	0.53	0.59	0.62	0.61	0.67	0.71	0.69	0.70
FIN	0.72	0.69	0.69	0.67	0.62	0.63	0.64	0.63	0.61	0.59	0.61	0.62
FRA	0.60	0.53	0.53	0.54	0.53	0.52	0.52	0.53	0.53	0.54	0.52	0.52
GBR	0.79	0.68	0.64	0.64	0.62	0.62	0.65	0.67	0.70	0.72	0.72	0.74
GRC	0.51	0.51	0.50	0.50	0.49	0.51	0.50	0.50	0.50	0.55	0.56	0.57
HUN	0.57	0.43	0.47	0.53	0.55	0.54	0.54	0.54	0.55	0.56	0.57	0.66

Table 3.8 (Continued)

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
IRL	0.75	0.63	0.59	0.56	0.54	0.49	0.65	0.71	0.77	**1.00**	**1.00**	**1.00**
ISL	0.83	0.73	0.70	0.70	0.77	0.76	0.78	0.76	0.77	0.78	0.80	0.72
ISR	0.61	0.56	0.58	0.64	0.68	0.67	0.68	0.76	0.70	0.70	0.76	0.77
ITA	0.49	0.49	0.50	0.52	0.51	0.51	0.52	0.51	0.52	0.54	0.57	0.58
JPN	0.87	1.00	0.78	0.78	0.74	0.74	0.69	0.72	0.76	0.77	0.76	0.77
KOR	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
LTU	0.66	0.67	0.60	0.61	0.56	0.57	0.59	0.66	0.72	0.75	0.79	0.80
LUX	0.71	0.81	0.69	0.71	0.67	0.66	0.71	0.72	0.76	0.78	0.74	0.75
LVA	0.68	0.67	0.65	0.66	0.60	0.58	0.62	0.62	0.64	0.69	0.72	0.74
NLD	0.78	0.70	**1.00**	0.74	0.67	0.67	0.71	0.72	0.75	0.76	0.76	0.79
NOR	0.66	0.61	0.58	0.65	0.57	0.57	0.59	0.62	0.64	0.62	0.55	0.55
NZL	0.75	0.63	0.63	0.72	0.71	0.69	0.65	0.80	0.81	0.83	0.82	0.83
POL	0.51	0.50	0.59	0.57	0.59	0.74	0.63	0.58	0.59	0.60	0.64	0.65
PRT	0.50	0.51	0.53	0.54	0.52	0.48	0.52	0.57	0.60	0.63	0.63	0.68
SVK	0.57	0.59	**1.00**	0.67	0.60	0.60	0.60	0.61	0.62	0.63	0.59	0.66
SVN	0.49	0.50	0.55	0.55	0.52	0.50	0.50	0.50	0.48	0.53	0.58	0.62
SWE	0.55	0.54	0.56	0.59	0.60	0.60	0.56	0.54	0.54	0.56	0.58	0.59
TUR	0.67	0.66	0.69	0.74	0.76	0.80	0.82	0.80	0.84	0.85	0.88	0.87
USA	0.92	0.81	0.79	0.75	0.71	0.72	0.73	0.80	0.97	0.81	0.80	0.84
Count	3	4	5	4	4	4	4	4	3	4	4	4
Average	0.71	0.68	0.68	0.67	0.65	0.65	0.66	0.68	0.70	0.71	0.71	0.73
Median	0.67	0.65	0.63	0.64	0.60	0.61	0.63	0.63	0.67	0.70	0.69	0.72
Min	0.49	0.43	0.47	0.50	0.49	0.48	0.50	0.50	0.48	0.53	0.52	0.52
Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Stdev	0.16	0.16	0.16	0.14	0.15	0.15	0.15	0.15	0.15	0.14	0.14	0.13

Source: Authors' calculations.

quite important for the country sample analysis. In addition, Table 3.9 provides a summary of the DEA results for the three models using an input-oriented assessment. The purpose of an input-oriented assessment is to study by how much input quantities can be proportionally reduced without changing the output quantities produced. The average efficiency score throughout the period is around 0.6 for the 1 input and 1 output model (Model 0) and around 0.7 in the alternative models (Models 1 and 2). Interestingly, the average input efficiency scores have increased slightly between 2006 and 2017. Nevertheless, these results imply that some possible efficiency gains could be achieved with around 30 per cent less government spending, on average, without changing the PSP outputs.

Figure 3.3 illustrates the production possibility frontier for the baseline model (Model 0), for 2006 (first year of our sample) and for 2017 (last year of our sample), pinpointing notably the countries that define the frontier: Switzerland and Korea in 2006, and Chile and Korea

Table 3.9 Summary of DEA results (input efficiency scores)

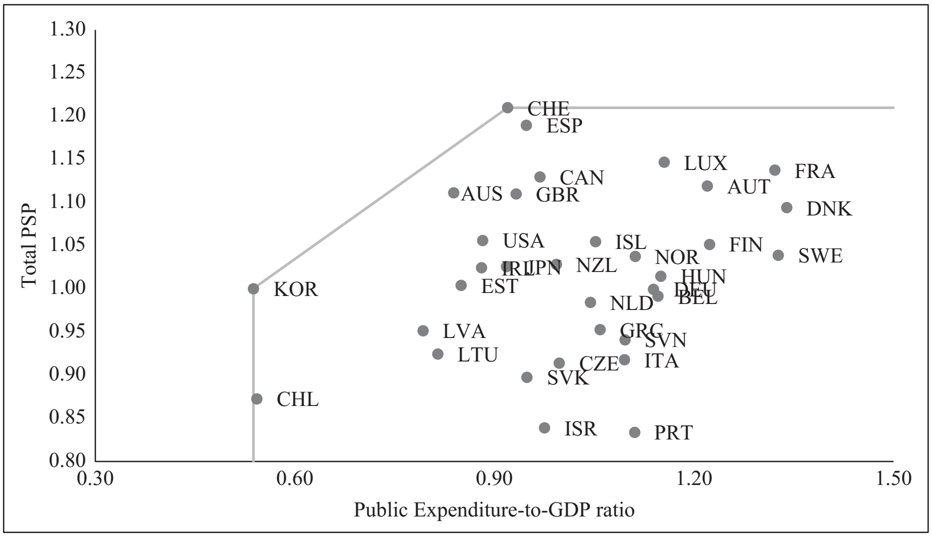
		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Model 0	Efficient	2	3	3	3	3	3	2	3	3	3	2	2
	Name	CHE; KOR	CHE; CHL; KOR	CHE; CHL; KOR	AUS; CHL; KOR	AUS; CHL; KOR	AUS; CHL; KOR	AUS; KOR	AUS; CHL; KOR	CHE; CHL; KOR	CHL; IRL; KOR	[illegible]	CHL; KOR
	Average	0.64	0.62	0.60	0.61	0.59	0.58	0.58	0.59	0.65	0.64	0.65	0.66
	Median	0.60	0.58	0.55	0.58	0.55	0.54	0.54	0.55	0.62	0.61	0.64	0.65
	Min	0.46	0.43	0.41	0.45	0.42	0.41	0.40	0.41	0.43	0.43	0.46	0.46
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Stdev	0.16	0.16	0.16	0.14	0.15	0.15	0.15	0.15	0.17	0.15	0.13	0.14
Model 1	Efficient	3	3	3	3	3	3	4	3	4	3	2	2
	Name	CHE; CHL; KOR	CHE; CHL; KOR	CHE; CHL; KOR	AUS; CHL; KOR	AUS; CHL; KOR	AUS; CHL; KOR	AUS; CHL; KOR; TUR	AUS; CHL; KOR	CHE; CHL; KOR; USA	CHL; IRL; KOR	CHL; KOR	CHL; KOR
	Average	0.70	0.67	0.65	0.68	0.66	0.66	0.66	0.68	0.73	0.72	0.73	0.73
	Median	0.66	0.63	0.60	0.65	0.61	0.63	0.63	0.66	0.70	0.70	0.71	0.73
	Min	0.52	0.47	0.47	0.54	0.51	0.50	0.49	0.48	0.50	0.52	0.53	0.50
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Stdev	0.15	0.15	0.14	0.13	0.13	0.13	0.13	0.13	0.15	0.13	0.12	0.13

Table 3.9 (Continued)

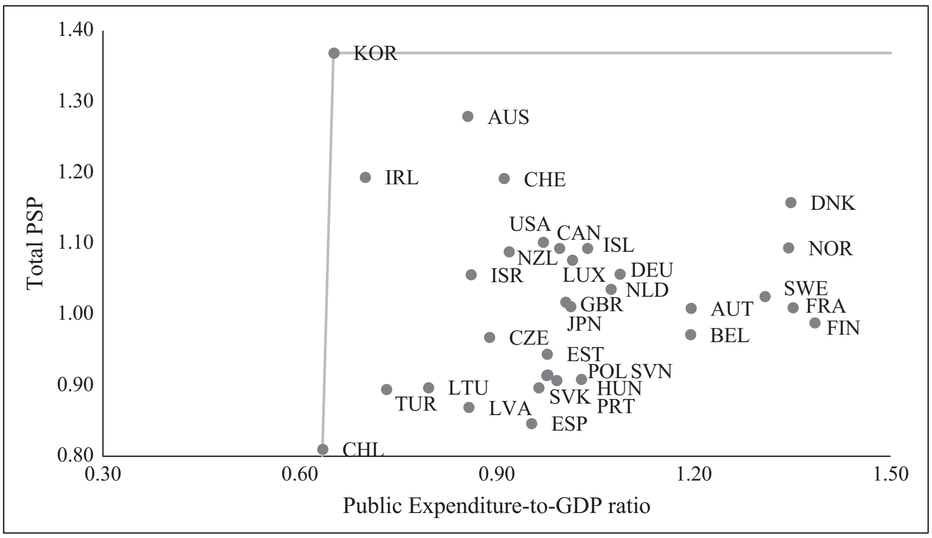
		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Model 2	Efficient	3	4	5	4	4	4	4	4	3	4	4	4
	Name	CHE; ESP; KOR	CHE; CHL; ESP; KOR	CHE; CHL; KOR; NLD; SVK	AUS; CHE; CHL; KOR	AUS; CHE; CHL; KOR	AUS; CHE; CHL; KOR	AUS; CHE; CHL; KOR	AUS; CHE; CHL; KOR	CHE; CHL; KOR	CHE; CHL; IRL; KOR	CHE; CHL; IRL; KOR	CHE; CHL; IRL; KOR
	Average	0.71	0.68	0.68	0.67	0.65	0.65	0.66	0.68	0.70	0.71	0.71	0.73
	Median	0.67	0.65	0.63	0.64	0.60	0.61	0.63	0.63	0.67	0.70	0.69	0.72
	Min	0.49	0.43	0.47	0.50	0.49	0.48	0.50	0.50	0.48	0.53	0.52	0.52
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Stdev	0.16	0.16	0.16	0.14	0.15	0.15	0.15	0.15	0.15	0.14	0.14	0.13

Source: Authors' calculations.

3.3.a –2006



3.3.b –2017



Note: Figure 3.3 plots the production possibility frontiers for Model 0 for the years 2006 and 2017. In the vertical axis we have the total Public Sector Performance (PSP) composite indicator AUS – Australia; AUT – Austria; BEL – Belgium; CAN – Canada; CHE – Switzerland; CHL – Chile; CZE – Czech Republic; DEU – Germany; DNK – Denmark; ESP – Spain; EST – Estonia; FIN – Finland; FRA – France; GBR – United kingdom; GRC – Greece; HUN – Hungary; IRL – Ireland; ISL – Iceland; ISR – Israel; ITA – Italy; JPN – Japan; KOR – South Korea; LTU – Lithuania; LUX – Luxembourg; LVA – Latvia; MEX – Mexico; NLD – Netherlands; NOR – Norway; NZL – New Zealand; POL – Poland; PRT – Portugal; SVK – Slovakia; SVN – Slovenia; SWE – Sweden; TUR – Turkey; USA – United States of America.

Source: Authors' calculations.

Figure 3.3 Production possibility frontier (input efficiency scores, Model 0) in 2017. For all the other countries inside the frontier, theoretically there would be room for improvement regarding efficiency gains.

Tables 3.10, 3.11 and 3.12 present the efficiency scores considering the output perspective. By computing output-oriented measures, one can assess how much output quantities can be proportionally increased without changing the input quantities used. Note that since Greece's PSP score is negative in 2012 and 2013, we cannot compute its efficiency score for Models 0 and 1.

Table 3.10 Output-oriented DEA VRS efficiency scores Model 0

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AUS	1.05	1.10	1.06	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	1.05	1.18	1.10	1.07
AUT	1.08	1.14	1.11	1.24	1.32	1.39	2.03	1.99	1.29	1.44	1.26	1.36
BEL	1.22	1.23	1.33	1.38	1.41	1.51	2.20	2.03	1.38	1.50	1.31	1.41
CAN	1.07	1.14	1.03	1.29	1.32	1.40	1.82	1.65	1.07	1.33	1.14	1.25
CHE	**1.00**	**1.00**	**1.00**	1.11	1.16	1.23	1.68	1.44	**1.00**	1.11	1.01	1.15
CHL	1.15	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
CZE	1.32	1.32	1.24	1.44	1.57	1.70	2.86	2.51	1.46	1.50	1.41	1.41
DEU	1.21	1.26	1.24	1.31	1.34	1.42	2.06	1.89	1.22	1.37	1.21	1.30
DNK	1.11	1.15	1.37	1.25	1.39	1.52	2.26	2.01	1.26	1.36	1.23	1.18
ESP	1.02	1.07	1.31	1.52	1.75	2.01	5.26	5.18	1.70	1.72	1.51	1.62
EST	1.17	1.24	3.40	1.43	1.82	1.68	2.13	2.25	1.42	1.60	1.38	1.45
FIN	1.15	1.15	1.29	1.26	1.38	1.47	2.59	2.19	1.41	1.58	1.33	1.39
FRA	1.06	1.17	1.34	1.37	1.44	1.51	2.25	1.99	1.33	1.48	1.31	1.36
GBR	1.09	1.15	1.52	1.33	1.45	1.60	2.12	1.89	1.20	1.34	1.20	1.35
GRC	1.27	1.29	1.64	1.56	2.37	3.69			2.13	2.25	1.91	1.92
HUN	1.19	1.62	1.65	1.63	1.92	2.04	3.61	2.54	1.55	1.73	1.60	1.50
IRL	1.16	1.21	2.09	1.48	1.67	1.68	2.69	2.07	1.20	**1.00**	1.21	1.15
ISL	1.15	1.10	1.18	1.29	1.72	1.63	2.39	1.91	1.34	1.29	1.09	1.25
ISR	1.44	1.39	1.23	1.34	1.20	1.19	1.45	1.25	1.29	1.46	1.21	1.30
ITA	1.32	1.38	1.84	1.61	1.77	2.00	4.95	3.50	1.81	1.97	1.77	1.71
JPN	1.18	1.03	1.52	1.36	1.39	1.63	2.08	1.76	1.34	1.39	1.24	1.35
KOR	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	1.03	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
LTU	1.25	1.27	1.39	1.50	1.94	1.81	2.27	2.23	1.37	1.52	1.36	1.53
LUX	1.06	1.06	1.37	1.21	1.22	1.37	2.20	1.59	1.12	1.32	1.19	1.27
LVA	1.20	1.27	3.04	1.56	2.56	1.97	2.49	2.54	1.69	1.62	1.39	1.58
NLD	1.23	1.13	1.00	1.10	1.24	1.35	2.30	2.14	1.34	1.42	1.29	1.32
NOR	1.17	1.17	1.27	1.19	1.32	1.40	1.66	1.62	1.15	1.22	1.21	1.25
NZL	1.18	1.20	1.47	1.24	1.27	1.36	1.61	1.25	1.11	1.29	1.03	1.26
POL	1.52	1.50	1.21	1.30	1.20	1.11	1.33	1.48	1.48	1.61	1.43	1.50

(Continued)

Table 3.10 (Continued)

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
PRT	1.45	1.43	1.39	1.58	1.66	2.08	6.12	3.66	1.70	1.79	1.57	1.51
SVK	1.35	1.38	1.09	1.46	1.51	1.71	2.27	2.31	1.54	1.57	1.43	1.53
SVN	1.29	1.31	1.14	1.37	1.63	1.83	3.98	2.89	1.50	1.71	1.49	1.51
SWE	1.16	1.18	1.47	1.28	1.23	1.40	2.25	1.82	1.27	1.23	1.13	1.33
TUR	1.43	1.57	1.68	1.58	1.44	1.33	1.45	1.32	1.18	1.32	1.38	1.53
USA	1.13	1.16	1.33	1.40	1.47	1.61	1.94	1.77	1.05	1.22	1.12	1.24
Count	2	3	3	3	3	3	2	3	3	3	2	2
Average	1.19	1.22	1.43	1.34	1.49	1.59	2.39	2.05	1.34	1.44	1.30	1.37
Median	1.17	1.18	1.33	1.34	1.41	1.51	2.20	1.95	1.33	1.42	1.26	1.35
Min	1	1	1	1	1	1	1	1	1	1	1	1
Max	1.52	1.62	3.40	1.63	2.56	3.69	6.12	5.18	2.13	2.25	1.91	1.92
Stdev	0.13	0.16	0.51	0.17	0.35	0.47	1.16	0.83	0.25	0.27	0.21	0.19

Source: Authors' calculations.

Table 3.11 Output-oriented DEA VRS efficiency scores Model 1

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AUS	1.02	1.07	1.03	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	1.02	1.12	1.10	1.07
AUT	1.08	1.14	1.11	1.24	1.32	1.39	2.03	1.99	1.29	1.44	1.26	1.36
BEL	1.22	1.23	1.33	1.38	1.41	1.51	2.20	2.03	1.38	1.50	1.31	1.41
CAN	1.07	1.14	1.03	1.29	1.32	1.40	1.82	1.65	1.07	1.33	1.14	1.25
CHE	**1.00**	**1.00**	**1.00**	1.10	1.14	1.18	1.64	1.44	**1.00**	1.07	1.01	1.15
CHL	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
CZE	1.32	1.32	1.24	1.44	1.57	1.70	2.86	2.51	1.46	1.50	1.41	1.41
DEU	1.21	1.26	1.24	1.31	1.34	1.42	2.06	1.89	1.22	1.37	1.21	1.30
DNK	1.11	1.15	1.37	1.25	1.39	1.52	2.26	2.01	1.26	1.36	1.23	1.18
ESP	1.02	1.07	1.31	1.52	1.75	2.01	5.26	5.18	1.64	1.72	1.51	1.62
EST	1.14	1.19	3.33	1.43	1.82	1.68	2.13	2.25	1.42	1.55	1.38	1.45
FIN	1.15	1.15	1.29	1.26	1.38	1.47	2.59	2.19	1.41	1.58	1.33	1.39
FRA	1.06	1.17	1.34	1.37	1.44	1.51	2.25	1.99	1.33	1.48	1.31	1.36
GBR	1.03	1.15	1.52	1.33	1.45	1.60	2.12	1.89	1.20	1.34	1.20	1.35
GRC	1.27	1.29	1.64	1.56	2.37	3.69			2.13	2.25	1.91	1.92
HUN	1.19	1.62	1.65	1.63	1.92	2.04	3.61	2.54	1.55	1.73	1.60	1.50
IRL	1.15	1.21	2.09	1.48	1.67	1.68	2.69	2.07	1.18	**1.00**	1.21	1.15
ISL	1.15	1.10	1.18	1.29	1.72	1.63	2.39	1.91	1.34	1.26	1.09	1.25
ISR	1.44	1.39	1.23	1.34	1.19	1.17	1.41	1.25	1.28	1.41	1.21	1.30
ITA	1.32	1.38	1.84	1.61	1.77	2.00	4.95	3.50	1.81	1.97	1.77	1.71

Table 3.11 (Continued)

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
JPN	1.17	1.03	1.52	1.33	1.39	1.62	2.08	1.76	1.34	1.39	1.24	1.35
KOR	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
LTU	1.24	1.25	1.39	1.50	1.94	1.81	2.27	2.18	1.35	1.51	1.36	1.53
LUX	1.06	1.06	1.37	1.21	1.22	1.37	2.20	1.59	1.10	1.32	1.19	1.27
LVA	1.19	1.24	2.94	1.53	2.56	1.96	2.49	2.54	1.69	1.57	1.39	1.58
NLD	1.23	1.13	1.00	1.10	1.24	1.35	2.30	2.14	1.34	1.42	1.29	1.32
NOR	1.17	1.17	1.27	1.19	1.32	1.40	1.66	1.62	1.15	1.22	1.21	1.25
NZL	1.17	1.20	1.47	1.24	1.27	1.36	1.61	1.25	1.11	1.23	1.03	1.26
POL	1.52	1.50	1.21	1.30	1.20	1.11	1.33	1.48	1.48	1.61	1.43	1.50
PRT	1.45	1.43	1.39	1.58	1.66	2.08	6.12	3.66	1.70	1.79	1.57	1.51
SVK	1.31	1.37	1.08	1.40	1.51	1.67	2.23	2.25	1.49	1.57	1.43	1.53
SVN	1.29	1.31	1.14	1.37	1.63	1.83	3.98	2.89	1.50	1.71	1.49	1.51
SWE	1.16	1.18	1.47	1.28	1.23	1.40	2.25	1.82	1.27	1.23	1.13	1.33
TUR	1.31	1.44	1.60	1.44	1.37	1.23	**1.00**	1.22	1.12	1.26	1.38	1.53
USA	1.04	1.07	1.28	1.35	1.47	1.61	1.94	1.77	**1.00**	1.09	1.12	1.24
Count	3	3	3	3	3	3	4	3	4	3	2	2
Average	1.18	1.21	1.43	1.33	1.49	1.58	2.37	2.04	1.33	1.43	1.30	1.37
Median	1.17	1.18	1.31	1.33	1.39	1.51	2.20	1.95	1.33	1.41	1.26	1.35
Min	1	1	1	1	1	1	1	1	1	1	1	1
Max	1.52	1.62	3.33	1.63	2.56	3.69	6.12	5.18	2.13	2.25	1.91	1.92
Stdev	0.13	0.15	0.50	0.17	0.35	0.47	1.18	0.83	0.26	0.28	0.21	0.19

Source: Authors' calculations.

Table 3.12 Output-oriented DEA VRS efficiency scores Model 2

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
AUS	1.01	1.05	1.05	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	1.02	1.08	1.09	1.01
AUT	1.05	1.06	1.05	1.05	1.06	1.07	1.08	1.08	1.09	1.11	1.12	1.14
BEL	1.13	1.13	1.13	1.14	1.15	1.14	1.14	1.14	1.14	1.16	1.14	1.17
CAN	1.07	1.08	1.03	1.08	1.07	1.08	1.08	1.09	1.06	1.11	1.12	1.11
CHE	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
CHL	1.01	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
CZE	1.28	1.31	1.21	1.30	1.29	1.28	1.28	1.31	1.31	1.30	1.31	1.27
DEU	1.08	1.08	1.08	1.09	1.10	1.10	1.09	1.09	1.10	1.10	1.12	1.12
DNK	1.03	1.03	1.05	1.06	1.06	1.05	1.09	1.11	1.09	1.09	1.10	1.06
ESP	**1.00**	**1.00**	1.24	1.25	1.23	1.22	1.20	1.19	1.21	1.23	1.24	1.25
EST	1.16	1.23	1.25	1.23	1.21	1.20	1.21	1.21	1.20	1.22	1.21	1.21

(*Continued*)

Table 3.12 (Continued)

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
FIN	1.02	1.02	1.02	1.04	1.04	1.03	1.02	1.02	1.03	1.05	1.04	1.04
FRA	1.06	1.09	1.10	1.11	1.10	1.10	1.11	1.11	1.13	1.14	1.16	1.17
GBR	1.09	1.12	1.16	1.17	1.15	1.12	1.11	1.11	1.12	1.12	1.12	1.14
GRC	1.26	1.25	1.38	1.39	1.40	1.41	1.42	1.39	1.38	1.41	1.43	1.44
HUN	1.18	1.38	1.44	1.42	1.37	1.37	1.37	1.36	1.35	1.40	1.46	1.36
IRL	1.13	1.19	1.21	1.19	1.19	1.16	1.11	1.11	1.11	**1.00**	**1.00**	**1.00**
ISL	1.02	1.05	1.06	1.05	1.05	1.06	1.07	1.09	1.10	1.10	1.08	1.11
ISR	1.20	1.22	1.23	1.27	1.19	1.16	1.17	1.15	1.29	1.28	1.19	1.17
ITA	1.31	1.35	1.46	1.43	1.41	1.41	1.41	1.34	1.36	1.41	1.40	1.40
JPN	1.06	1.00	1.09	1.10	1.10	1.09	1.10	1.08	1.07	1.07	1.09	1.10
KOR	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
LTU	1.24	1.21	1.38	1.36	1.36	1.35	1.33	1.31	1.27	1.24	1.25	1.24
LUX	1.05	1.03	1.11	1.08	1.07	1.08	1.08	1.07	1.07	1.09	1.11	1.11
LVA	1.18	1.18	1.43	1.39	1.38	1.39	1.37	1.34	1.32	1.32	1.36	1.35
NLD	1.06	1.06	**1.00**	1.03	1.07	1.06	1.04	1.04	1.04	1.03	1.05	1.05
NOR	1.11	1.12	1.15	1.10	1.14	1.15	1.11	1.12	1.12	1.11	1.14	1.11
NZL	1.10	1.13	1.15	1.10	1.10	1.09	1.06	1.02	1.07	1.08	1.02	1.10
POL	1.47	1.49	1.10	1.25	1.13	1.03	1.27	1.30	1.38	1.38	1.39	1.34
PRT	1.23	1.23	1.25	1.23	1.23	1.22	1.20	1.19	1.18	1.20	1.23	1.21
SVK	1.34	1.36	**1.00**	1.36	1.38	1.43	1.42	1.46	1.45	1.42	1.41	1.38
SVN	1.29	1.29	1.11	1.23	1.25	1.28	1.27	1.29	1.30	1.32	1.33	1.32
SWE	1.11	1.09	1.09	1.07	1.06	1.07	1.10	1.10	1.13	1.12	1.11	1.11
TUR	1.42	1.43	1.48	1.41	1.30	1.23	1.20	1.19	1.17	1.22	1.23	1.24
USA	1.04	1.08	1.09	1.10	1.11	1.12	1.12	1.11	1.02	1.09	1.09	1.07
Count	3	4	5	4	4	4	4	4	3	4	4	4
Average	1.14	1.15	1.16	1.17	1.16	1.16	1.16	1.16	1.16	1.17	1.18	1.17
Median	1.10	1.12	1.11	1.11	1.13	1.12	1.11	1.11	1.12	1.12	1.12	1.14
Min	1	1	1	1	1	1	1	1	1	1	1	1
Max	1.47	1.49	1.48	1.43	1.41	1.43	1.42	1.46	1.45	1.42	1.46	1.44
Stdev	0.13	0.14	0.15	0.14	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.13

Source: Authors' calculations.

Finally, Table 3.13 provides a summary of the DEA results for the three models using output oriented models. The average output efficiency score is approximately 1.50 for Models 0 and 1 and 1.16 for Model 3, suggesting that outputs could be increased by approximately 50 per cent or 16 per cent. The output efficiency scores for Models 0 and 1 where somewhat higher and seemed to have peaked in the period 2011–13, and then they decreased.

Table 3.13 *Summary of DEA results (output efficiency scores)*

		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Model 0	Efficient	2	3	3	3	3	3	2	3	3	3	2	2
	Name	CHE; KOR	CHE; CHL; KOR	CHE; CHL; KOR	AUS; CHL; KOR	AUS; CHL; KOR	AUS; CHL; KOR	AUS; KOR	AUS; CHL; KOR	CHE; CHL; KOR	CHL; IRL; KOR	CHL; KOR	CHL; KOR
	Average	1.19	1.22	1.43	1.34	1.49	1.59	2.39	2.05	1.34	1.44	1.30	1.37
	Median	1.17	1.18	1.33	1.34	1.41	1.51	2.20	1.95	1.33	1.42	1.26	1.35
	Min	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Max	1.52	1.62	3.40	1.63	2.56	3.69	6.12	5.18	2.13	2.25	1.91	1.92
	Stdev	0.13	0.16	0.51	0.17	0.35	0.47	1.16	0.83	0.25	0.27	0.21	0.19
Model 1	Efficient	3	3	3	3	3	3	4	3	4	3	2	2
	Name	CHE; CHL; KOR	CHE; CHL; KOR	CHE; CHL; KOR	AUS; CHL; KOR	AUS; CHL; KOR	AUS; CHL; KOR	AUS; CHL; KOR; TUR	AUS; CHL; KOR	CHE; CHL; KOR; USA	CHL; IRL; KOR	CHL; KOR	CHL; KOR
	Average	1.18	1.21	1.43	1.33	1.49	1.58	2.37	2.04	1.33	1.43	1.30	1.37
	Median	1.17	1.18	1.31	1.33	1.39	1.51	2.20	1.95	1.33	1.41	1.26	1.35
	Min	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Max	1.52	1.62	3.33	1.63	2.56	3.69	6.12	5.18	2.13	2.25	1.91	1.92
	Stdev	0.13	0.15	0.50	0.17	0.35	0.47	1.18	0.83	0.26	0.28	0.21	0.19

(*Continued*)

Table 3.13 (Continued)

		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Model 2	Efficient	3	4	5	4	4	4	4	4	3	4	4	4
	Name	CHE; ESP; KOR	CHE; CHL; ESP; KOR	CHE; CHL; KOR; NLD; SVK	AUS; CHE; CHL; KOR	AUS; CHE; CHL; KOR	AUS; CHE; CHL; KOR	AUS; CHE; CHL; KOR	AUS; CHE; CHL; KOR	CHE; CHL; KOR	CHE; CHL; IRL; KOR	CHE; CHL; IRL; KOR	CHE; CHL; IRL; KOR
	Average	1.14	1.15	1.16	1.17	1.16	1.16	1.16	1.16	1.16	1.17	1.18	1.17
	Median	1.10	1.12	1.11	1.11	1.13	1.12	1.11	1.11	1.12	1.12	1.12	1.14
	Min	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Max	1.47	1.49	1.48	1.43	1.41	1.43	1.42	1.46	1.45	1.42	1.46	1.44
	Stdev	0.13	0.14	0.15	0.14	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.13

Source: Authors' calculations.

6. CONCLUSION

In this study, we provided a review of the literature dealing with overall public sector performance and efficiency. Moreover, we outlined a methodology to assess public sector efficiency and we have created a novel and large cross-country panel dataset of government indicators and public sector efficiency scores, covering all 36 OECD countries over the 2006–17 time period. In practice, we used economic and sociodemographic indicators to construct performance composite indicators, and then we computed input and output oriented efficiency scores solving the several DEA problems.

The average input efficiency score in the period 2006–17 was found to be around 0.6–0.7, implying that some efficiency gains could be achieved with government expenditure of approximately 30–40 per cent less, on average, without changing the overall level of performance. The average output efficiency score was found to be between 1.16 and 1.50, suggesting that outputs could be increased by approximately 16–50 per cent.

With this study, we have filled a gap in the literature, by providing a cross-country dataset of public sector performance indicators and efficiency scores, which can be useful for further research by other authors.

NOTES

1. The analysis of government size with respect to economic growth has recently received more attention in the context of empirical analysis. The existence of a relationship between the variables was firstly postulated by the German political economist Adolph Wagner (1911). Lamartina and Zaghini (2011) provided empirical evidence for a positive relationship between government size and GDP per capita using panel of 23 OECD countries.
2. The government intervenes in the economy in four ways (Labonte, 2010). First, it produces goods and services, such as infrastructure, education and national defence. Second, it transfers income, both vertically across income levels and horizontally among groups with similar incomes and different characteristics. Third, it taxes to pay for its outlays, which can lower economic efficiency by distorting behaviour. Finally, government regulation alters economic activity.
3. The Web of Science was chosen as it represents one of the major academic search engines in social sciences and facilitates a wide-ranging identification of relevant publications.
4. Within the public sector literature, some studies have evaluated government performance of a specific government function or the performance of local governments. In terms of local governance performance, see for instance Van den Eeckaut et al. (1993), De Borger et al. (1994) and De Borger and Kerstens (1996, 2000) for Belgium; Athanassopoulos and Triantis (1998) and Doumpos and Cohen (2014) for Greece; Worthington (2000) for Australia; Prieto and Zofio (2001), Balaguer-Coll et al. (2002) and Benito et al. (2010) for Spain; Storto (2015) for Italy; Waldo (2001) for Sweden; and Sampaio and Stosic (2005) for Brazil. In Portugal, we highlight the studies of Afonso and Fernandes (2006, 2008), Afonso and Scaglioni (2007), Cruz and Marques (2014) and Afonso and Venâncio (2016).
5. 'The need to improve competitiveness, concerns about fiscal sustainability and growing demands by taxpayers to get more value for public money as well as the need to reconsider the scope for state intervention in the economy has prompted efforts to increase the focus of budgets on more growth-enhancing activities and gear the tax mix and the allocation of resources within the public sector towards better efficiency and effectiveness' (EC, 2007, p. 9).
6. For instance, Ruggiero (2004) and Simar and Wilson (2007) provide an overview of this issue.
7. The 36 OECD countries considered are: Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States.
8. For Chile, Iceland, Israel, South Korea and Mexico, we use the data available in Medina and Schneider (2017).
9. For Switzerland, we were only able to collect data for the period between 2009 and 2016.
10. From IMF World Economic Outlook (WEO database), we retrieved data for Greece for the period between 2006 and 2012 and for the United States for the periods 2005 and 2007.

11. We were not able to collect data on the following countries: Australia, Canada, Mexico, New Zealand, Chile, Israel and South Korea.
12. From IMF World Economic Outlook (WEO database), we retrieved data for New Zealand for the periods 2005 and 2012. For Turkey, we retrieve data from European Commission, AMECO database. For Chile and Iceland we were only able to collect data for the period between 2013 and 2016. For Turkey, we were only able to get data for the period between 2009 and 2015. We were not able to collect data for Canada.
13. From IMF World Economic Outlook (WEO database), we retrieved data for Canada for the period between 2005 and 2012 and for New Zealand for the period 2009 and 2012. For Turkey, we retrieve data from European Commission, AMECO database. We were not able to collect data for Mexico. For Chile and Iceland, we were only able to collect data for the period between 2013 and 2016. For New Zealand, we were only able to collect data for the period between 2009 and 2016. For Japan, we were only able to collect data for the period between 2005 and 2016.
14. DEA is a non-parametric frontier methodology, which draws from Farrell's (1957) seminal work and that was further developed by Charnes et al. (1978). Coelli et al. (2002) and Thanassoulis (2001) offer introductions to DEA.
15. When a DMU is efficient by default it means that it will not appear as peer of any other non-efficient DMU.

REFERENCES

Adam, A., Delis, M., Kammas, P. (2011). 'Public sector efficiency: leveling the playing field between OECD countries', *Public Choice*, 146, 163–183.

Afonso, A., Kazemi, M. (2017). 'Assessing Public Spending Efficiency in 20 OECD Countries', in Inequality and Finance in Macrodynamics (Dynamic Modeling and Econometrics in Economics and Finance), Bettina Bökemeier and Alfred Greiner (Editors). Springer.

Afonso, A., Romero, A., Monsalve, E. (2013). 'Public Sector Efficiency: Evidence for Latin America Public'. Inter-American Development Bank, 80478.

Afonso, A., St. Aubyn, M. (2005). 'Cross-country Efficiency of Secondary Education Provision: a Semi-parametric Analysis with Non-discretionary Inputs', European Central Bank, Working Paper n. 494.

Afonso, A., St. Aubyn, M. (2006). 'Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs', *Economic Modelling*, 23 (3), 476–491.

Afonso, A., Schuknecht, L. (2019). How 'Big' Should Government Be? (March 30, 2019). REM Working Paper 078-2019, Available at SSRN: https://ssrn.com/abstract=3362749

Afonso, A., Schuknecht, L., Tanzi, V. (2005). 'Public sector efficiency: An international comparison', *Public Choice*, 123 (3–4), 321–347.

Afonso, A., Venâncio, A. (2019). 'Local Territorial Reform and Regional Spending Efficiency', Local Government Studies, forthcoming.

Antonelli, M., de Bonis, V. (2019). 'The efficiency of social public expenditure in European countries: a two-stage analysis', *Applied Economics*, 51 (1), 47–60.

Chan, S. -G., Ramly, Z., Karim, M. (2017). 'Government spending efficiency on economic growth: Roles of value-added tax', *Global Economic Review*, 46 (2), 162–188.

Dutu, R., Sicari, P. (2016). 'Public Spending Efficiency in the OECD: Benchmarking Health Care, Education and General Administration', OECD Economics Department Working Papers 1278.

EC (2007). The EU Economy: 2007 Review, Moving Europe's Productivity Frontier, November, European Commission.

Gupta, S., Verhoeven, M. (2001). 'The efficiency of government expenditure. Experiences from Africa', *Journal of Policy Modelling*, 23, 433– 467.

Herrera, S., Ouedraogo, A. (2018). Efficiency of Public Spending in Education, Health, and Infrastructure: An International Benchmarking Exercise, World Bank Policy Research Working Paper 8586.

Labonte, M. (2010), 'The Size and Role of Government: Economic Issues', Congressional Research Service 7-5700, RL32162.

Lamartina, S., Zaghini, A. (2011), 'Increasing public expenditure: Wagner's law in OECD countries', *German Economic Review*, 12 (2), 149–164.

Mandl, U., Dierx, A., Ilzkovitz, F. (2008). The effectiveness and efficiency of public spending. *European Economy. Economic Papers*, 301, 1−34.

Medina, L., Schneider, F. (2018). Shadow Economies Around the World: What Did We Learn Over the Last 20 Years? IMF Working Papers. 18.

Montes, G., Bastos, J., de Oliveira, A. (2019). 'Fiscal transparency, government effectiveness and government spending efficiency: Some international evidence based on panel data approach', *Economic Modelling*, 79, 211–225.

Mohanty, R., Bhanumurthy, N. (2018). 'Assessing Public Expenditure Efficiency at Indian States', National Institute of Public Finance and Policy, New Delhi, NIPFP Working Paper 225.

Ruggiero, J. (2004). 'Performance evaluation when nondiscretionary factors correlate with technical efficiency'. *European Journal of Operational Research*, 159, 250–257.

Schneider, F. (2016). 'Comment on Feige's paper, reflections on the meaning and measurement of unobserved economies: What do we really know about the "shadow economy?"', *Journal of Tax Administration*, 2 (2), 82–92.

Simar, L., Wilson, P. (2007). 'Estimation and inference in two-stage, semi-parametric models of production processes', *Journal of Econometrics*, 136 (1), 31–64.

Tanzi, V., Schuknecht, L. (1997). 'Reconsidering the fiscal role of government: The international perspective', *American Economic Review*, 87 (2), 164−168.

Tanzi, V., Schuknecht, L. (2000). *Public Spending in the 20th Century: A Global Perspective*. Cambridge: Cambridge University Press.

Wagner, A. (1911). Staat in nationalökonomischer Hinsicht. In Handwörterbuch der Staatswissenschaften. Third edition, Book VII. Jena: Lexis. 743–745.

4. Methodologies for assessing government efficiency

Caitlin O'Loughlin, Léopold Simar, and Paul W. Wilson

1. INTRODUCTION

Microeconomic theory of the firm is concerned with the ability of firms to transform inputs such as labor, physical capital, energy and materials into outputs that might include various goods or services depending on a particular industry. Standard graduate-level textbook treatments (for example, Henderson and Quandt, 1971; Varian, 1978; or Mas-Colell et al., 1995) posit cost-minimizing and profit-maximizing behavior among firms and typically assume firms do these things optimally, although some (for example, Lesourne, 1972) consider public-sector institutions. But the theory of production is much larger. The theory of production and efficiency analysis examines how production units (which might be for-profit firms, non-profit organizations, government entities such as agencies or local governments, or other decision-making units such as hospitals, schools and so on) transform their inputs into outputs, and in addition *how well* they do so. Efficiency analyses have been developed to examine the performance of not only for-profit firms, but also decision-making units that might not explicitly have maximization as a goal.

Production theory (as opposed to microeconomic theory of the firm) builds on the pioneering work of Koopmans (1951) and Debreu (1951). Farrell (1957) and Afriat (1972) were among the first to look at the empirical issues of how to estimate production efficiency. The efficient production frontier is defined in an appropriate input–output space as the locus of optimal combinations of input and output quantities. One may consider either the locus of the maximal attainable level of outputs corresponding to a given level of inputs, or when output quantities are regarded as exogenous, the researcher may search for the minimal levels of inputs necessary for producing a given amount of outputs. We will see below that more general situations can be also analyzed.

In this chapter, we discuss and illustrate modern, state-of-the-art methods that can be used to analyze and quantify in particular the productive performance of local, municipal governments. Municipal governments provide varying bundles of goods, services and amenities for residents, who in turn are free to choose among municipalities and hence the varying offerings. In addition to police and fire protection, municipal governments may provide roads and streets, traffic management, trash collection, street cleaning, water services, libraries and other services. In principle, municipalities compete with each other both in terms of taxation as well as provision of services along the lines of Tiebout (1956) and Ellickson (1971). Moreover, while municipal governments offer varying bundles of goods and services, they also vary in terms of *how well* or how efficiently they provide their particular bundles, that is, in terms of levels of waste that are incurred.

A number of studies have examined the production performance of municipal governments. For example, Grosskopf et al. (2001) suggest that competition among municipalities may create incentives to provide services efficiently by influencing citizens' willingness to pay for public services or their inclination to remain in the jurisdiction. Resident satisfaction over time also likely plays a role in this decision; Tran and Dollery (2020) note a strong relationship with this and municipal efficiency across different categories of municipal governments. Hayes et al. (1998), Grossman et al. (1999) and others find empirical evidence that competition among local governments tends to enhance efficiency. At the same time, friction caused by real estate transaction fees, costs of commuting and job search, and other factors may reduce competition among municipalities, perhaps leading to inefficient provision of services.

While there is a substantial literature on municipal efficiency, most studies offer little or no justification for the choice of the estimators used, and few if any provide statistical inference, presenting instead only point estimates with no accompanying measure or indication of the surrounding uncertainty. Examples can be found in many of the studies reviewed by Tang (1997), De Borger and Kerstens (2000), Afonso (2008), Da Cruz and Marques (2014), de Oliveira Junqueira (2015) and Narbón-Perpiñá and De Witte (2018). Empirical analyses of efficiency among municipal governments can be broadly divided into those that employ fully parametric methods along the lines of Aigner et al. (1977) and Meeusen and van den Broeck (1977) versus those that use fully nonparametric methods such as the data envelopment analysis (DEA) estimators proposed by Farrell (1957) and popularized by Banker et al. (1984) or free disposal hull (FDH) estimators proposed by Deprins et al. (1984). The focus in this chapter is on nonparametric methods, which are more often used to analyze municipal efficiency than parametric methods. Among studies using nonparametric methods, DEA estimators are used much more often than FDH estimators to analyze production by local governments.[1] As seen below, the choice between FDH and DEA estimators is not innocuous.

Before efficiency or any other measure of productive performance can be estimated, a well-specified model of the production process is needed. Specifically, a well-specified statistical model is needed, for without such a model, one cannot know what is estimated nor is inference possible. Only within the context of a properly specified model can the 'efficiency scores' of individual local governments be characterized. As seen below, these efficiency scores will be based on the distance from a particular municipality's input–output combination to the efficient frontier in a specific direction in the input–output space. The nonparametric model presented below involves minimal assumptions, and relies only on mild assumptions suggested by economic theory and, except for some additional mild regularity conditions, no parametric restrictions are imposed on the stochastic components of the model. As opposed to parametric models, the Data Generating Process (DGP) that specifies how observed data are generated in the model used below relies on only a few flexible assumptions. We describe recent theoretical developments that provide tools for making inference about efficiency as well as features of the model, including whether the production set is convex, returns to scale, the role of environmental variables and evolution of the frontier over time, as well as differences in mean efficiency across groups of municipalities.

The remaining part of this chapter is organized along the following lines. In the next section we define the basic model, concepts and notation for the production process. We then introduce some popular nonparametric estimators of the frontier and of efficiency, including the estimation of the efficiency scores. We summarize the main statistical properties of these estimators as well as estimation of confidence intervals for efficiency of individual decision-making units and recent Central Limit Theorem (CLT) results developed for averages of efficiency scores. In Section 4 we show how the CLT results can be used to build test statistics for different hypotheses about model features and how these might change over time. Section 7 illustrates how the methods discussed in Sections 2–6 can be used to analyze the performance of local, municipal governments in the United States. Conclusions are given in Section 8 with suggestions for some extensions.

2. A NONPARAMETRIC MODEL OF PRODUCTION

To establish notation and basic ideas, consider the vector $x \in \mathbb{R}_+^p$ of input quantities used to produce the output quantities $y \in \mathbb{R}_+^q$. The *production set*, that is, the set of feasible combinations of input and output quantities is given by

$$\Psi = \{(x, y) \mid x \text{ can produce } y\}. \tag{4.1}$$

Of particular interest is the *technology*, or efficient boundary of Ψ defined as

$$\Psi^\partial = \{(x, y) \in \Psi \mid (\gamma^{-1}x, \gamma y) \notin \Psi \text{ for all } \gamma > 1\}. \tag{4.2}$$

We assume the usual characteristics for Ψ coming from microeconomic theory of the firm; that is, Ψ is closed and both inputs and outputs are freely disposable.[2] We rule out free lunches with the assumption that production of any non-zero output requires use of a non-zero quantity of at least one input. Assuming free disposability of inputs and outputs amounts to assuming that the frontier is weakly monotonic, that is, non-decreasing in inputs.

The efficiency of a production plan (x, y) is measured by its distance to the efficient boundary. This measure depends on the chosen direction to the frontier. The Debreu-Farrell input efficiency measure

$$\theta(x, y \mid \Psi) = \inf\{\theta > 0 \mid (\theta x, y) \in \Psi\} \tag{4.3}$$

is the radial distance of (x, y) to its projection on Ψ^∂ along the ray x. In other words, $\theta(x, y \mid \Psi)$ is the smallest factor by which input quantities can be proportionately reduced while maintaining production at output levels y. Alternatively, the Debreu-Farrell output efficiency measure

$$\lambda(x, y \mid \Psi) = \sup\{\lambda > 0 \mid (x, \lambda y) \in \Psi\} \tag{4.4}$$

gives the maximum feasible, radial expansion of the outputs to reach the efficient frontier. Efficiency can also be measured in other directions. For example, the hyperbolic measure

$$\gamma(x, y \mid \Psi) = \sup\{\gamma > 0 \mid (\gamma^{-1}x, \gamma y) \in \Psi\} \tag{4.5}$$

suggested by Färe et al. (1985) gives the simultaneous, feasible proportionate reduction of input quantities and proportionate, feasible increase of output quantities for a unit operating at $(x, y) \in \Psi$ to reach the efficient frontier.

In most situations Ψ, and hence $\theta(x, y \mid \Psi)$, $\lambda(x, y \mid \Psi)$ and $\gamma(x, y \mid \Psi)$, are unknown and not observable. In empirical applications, the task at hand is to estimate these from a sample $\chi_n = \{(X_i, Y_i)\}_{i=1}^n$ of observations on input and output quantities of individual production units. To understand what can be learned from the observations in χ_n, we need a statistical model that specifies the DGP, that is, a set of assumptions that characterize how the observations are generated. As noted in Section 1, we focus here on nonparametric models requiring a minimal set of regularity assumptions.

To define the DGP, we assume the n observations (X_i, Y_i) are identically, independently distributed (iid) realizations of a random variables (X, Y) having Ψ as support. In other words, all observations lie within the production set, Ψ. Hence the joint density $f_{XY}(x, y)$ of (X, Y) is assumed to be zero everywhere outside Ψ. In addition, we assume this joint density is strictly positive along the frontier (that is, $f_{XY}(x, y) > 0 \; \forall \; (x, y) \in \Psi^{\partial}$), and continuously differentiable close to Ψ^{∂} in the interior of Ψ so that there is not a probability mass along the frontier. To establish properties of DEA and FDH estimators, the frontier must also be sufficiently smooth. Assumptions regarding convexity of Ψ or constant versus variable returns to scale may be made as appropriate. For technical details and precise statements of the various assumptions needed, see Kneip et al. (2008, 2015b).

3. NONPARAMETRIC ESTIMATION AND INFERENCE

3.1 Nonparametric Efficiency Estimators

Nonparametric estimation of the efficiency measures given in (4.3)–(4.5) requires replacing the unknown Ψ in the definitions of the measures with a suitable nonparametric estimator of Ψ. The most popular nonparametric estimators of Ψ based on the sample χ_n are the *envelopment* estimators that envelop the sample observations in a minimal (subject to some conditions) set spanning the observed points $\{(X_i, Y_i)\}_{i=1}^n$. The FDH estimator, introduced by Deprins et al. (1984) relies only on the assumption of free disposability of inputs and outputs, and thus is the least-restrictive, most-flexible of the envelopment estimators. The FDH estimator is given by

$$\widehat{\Psi}_{\text{FDH},n} = \{(x, y) \in \mathbb{R}_+^{p+q} \mid y \le Y_i, x \ge X_i, (X_i, Y_i) \in \chi_n\}, \tag{4.6}$$

that is, the union of the positive orthants in the inputs and the negative orthants in the outputs, whose origin coincides with the data points in χ_n. Since under the assumptions of the model outlined in Section 2 all observations (X_i, Y_i) must lie in Ψ, necessarily $\widehat{\Psi}_{\text{FDH},n} \subseteq \Psi$. In this sense, $\widehat{\Psi}_{\text{FDH},n}$ is a biased estimator of Ψ, as it lies in the interior of Ψ and the intersection of $\widehat{\Psi}_{\text{FDH},n}$ and the complement of Ψ is the null set.

Alternatively, if one is willing to assume Ψ is a convex set, then the variable returns to scale (VRS) version of the DEA estimator can be used to estimate Ψ. The VRS-DEA estimator of Ψ is the convex hull of $\widehat{\Psi}_{\text{FDH},n}$, that is, the smallest convex set containing $\widehat{\Psi}_{\text{FDH},n}$, and can be written as

$$\widehat{\Psi}_{\text{VRS},n} = \{(x, y) \in \mathbb{R}_+^{p+q} \mid y \le \sum_{i=1}^n \omega_i Y_i, x \ge \sum_{i=1}^n \omega_i X_i, \sum_{i=1}^n \omega_i = 1, \omega_i \ge 0, i = 1, \dots, n\}. \tag{4.7}$$

If, in addition to convexity of Ψ, one is willing to assume constant returns to scale (CRS), then the CRS version of the DEA estimator can be used to estimate Ψ. The CRS-DEA estimator of Ψ is given by

$$\widehat{\Psi}_{\text{CRS},n} = \{(x, y) \in \mathbb{R}_+^{p+q} \mid y \le \sum_{i=1}^n \omega_i Y_i, x \ge \sum_{i=1}^n \omega_i X_i, \omega_i \ge 0, i = 1, \dots, n\}. \tag{4.8}$$

This is the conical hull of $\widehat{\Psi}_{\mathrm{VRS},n}$ with vertex at the origin. By construction, when Ψ is convex and CRS holds, $\widehat{\Psi}_{\mathrm{FDH},n} \subseteq \widehat{\Psi}_{\mathrm{VRS},n} \subseteq \widehat{\Psi}_{\mathrm{CRS},n} \subseteq \Psi$. If Ψ is convex and CRS does not hold, then necessarily $\widehat{\Psi}_{\mathrm{FDH},n} \subseteq \widehat{\Psi}_{\mathrm{VRS},n} \subseteq \Psi$ and $\widehat{\Psi}_{\mathrm{CRS},n} \cap \overline{\Psi} \neq \varnothing$ where $\overline{\Psi}$ denotes the complement of Ψ. Both the VRS-DEA and CRS-DEA estimators were discussed by Farrell (1957), but received little attention until they were revisited and popularized by Charnes et al. (1978) and Banker et al. (1984).

Estimators of the efficiency of a particular production plan (x, y) are obtained by replacing Ψ in (4.3), (4.4) or (4.5) with one of the estimators of Ψ described above. For example, the input-oriented FDH efficiency estimator is $\hat{\theta}_{\mathrm{FDH},n}(x, y) = \theta(x, y \mid \widehat{\Psi}_{\mathrm{FDH},n})$. This is easily computed as

$$\hat{\theta}_{\mathrm{FDH},n}(x, y) = \min_{i \in D_{x,y}} \max_{j=1,\dots,p} \left(\frac{X_i^{(j)}}{x^{(j)}} \right), \tag{4.9}$$

where $D_{x,y} = \{i \mid (X_i, Y_i) \in \chi_n, X_i \leq x, Y_i \geq y\}$ is the set of indices of observations in χ_n dominating the point (x, y) and $a^{(j)}$ denotes the jth component of a vector a. The DEA-VRS is obtained as the solution to the linear program

$$\hat{\theta}_{\mathrm{VRS},n}(x, y) = \min_{\theta,\omega} \{\theta \mid y \leq \sum_{i=1}^{n} \omega_i Y_i, \theta x \geq \sum_{i=1}^{n} \omega_i X_i, \sum_{i=1}^{n} \omega_i = 1; \omega_i \geq 0 \ \forall\ i = 1, \dots, n\}. \tag{4.10}$$

The CRS-DEA estimator $\hat{\theta}_{\mathrm{CRS},n}(x, y)$ is obtained by dropping the constraint $\Sigma_{i=1}^{n} \omega_i = 1$ in (4.10). For computational details and other variations, see the detailed review in Simar and Wilson (2013).

3.2 Estimation and Confidence Intervals for Individual Efficiencies

The statistical properties of the nonparametric estimators defined above have been derived in a number of papers, including Kneip et al. (1998), Park et al. (2000), Kneip et al. (2008, (2015b) and (2016)), Park et al. (2010), Wilson (2011) and Daouia et al. (2017). See also Simar and Wilson (2015) and the references therein for a comprehensive review. To summarize, for a fixed point of interest (x, y) we have under mild regularity assumptions (see above)

$$n^{\kappa} \left(\hat{\theta}_{\bullet,n}(x, y) - \theta(x, y \mid \Psi) \right) \xrightarrow{\mathcal{L}} Q_{\bullet}(\eta_{x,y}) \tag{4.11}$$

as $n \to \infty$, where κ determines the rate of convergence and depends on the assumptions and the chosen estimator with '$\bullet$' representing FDH, VRS or CRS, and $Q_{\bullet}$ is a non-degenerate distribution depending on unknown parameters $\eta_{x,y}$. Hence, the various nonparametric estimators that have been discussed are statistically consistent, with known limiting distributions. Moreover, the rates of convergence are also known. The convergence rate n^{κ} determines the order of the error of estimation: we have $(\hat{\theta}_{\bullet,n}(x,y) - \theta(x,y \mid \Psi)) = O_p(n^{-\kappa})$. The error of estimation becomes smaller in probability terms as the sample size n increases. The larger is κ, the faster the error decreases as n increases. Similar results hold for the output-oriented and hyperbolic efficiency estimators. In particular, neither the rate of convergence n^{κ} nor the form of the limiting distribution $Q_{\bullet}(\eta_{x,y})$ depend on the direction from the fixed point of interest (x, y) to the frontier, although the constants in $\eta_{x,y}$ may differ.

For the FDH estimator, provided free disposability of inputs and outputs holds, $\kappa = 1/(p + q)$. Under free disposability, convexity of Ψ, and non-CRS, the VRS-DEA estimator has the rate n^{κ} with $\kappa = 2/(p + q + 1)$. With the additional assumption of CRS, the CRS-DEA estimator (as well as the VRS-DEA estimator, as proved by Kneip et al., 2016) converges at rate n^{κ} with $\kappa = 2/(p + q)$. The convergence rates illustrate the price of using nonparametric estimators; whereas parametric estimators typically converge at rate $n^{1/2}$, here the convergence rate of our nonparametric estimators becomes slower as the number $(p + q)$ of inputs and outputs increases. This is typical for nonparametric estimators, and is called the *curse of dimensionality*.[3] However, obtaining the typical parametric, root-n rate depends crucially on having the correct model specification, which may be unlikely. If a parametric model is mis-specified, then convergence may be rather fast, but the parametric estimators converge to some unknown thing that is not a feature of the underlying, true model. Robinson (1988) refers to this as 'root-n *in*consistency.' Consistency is arguably the most fundamental, important property of an estimator; if an estimator is consistent, then additional data will improve the estimate in probabilistic terms. If an estimator is inconsistent, then additional data may not be informative.

Another complication for the nonparametric efficiency estimators arises from the fact that the limiting distributions $Q_{\bullet}(\eta_{x,y})$ in ([illegible]) and similar expressions for the other estimators depend on unknown parameters in $\eta_{x,y}$ that are difficult to estimate. In the case of the FDH estimator, $Q_{\text{FDH}}(\eta_{x,y})$ is a Weibull distribution as shown in Park et al. (2000), but for the VRS-DEA case (Kneip et al., 2008) and for the CRS-DEA case (Park et al., 2010), the existence of the limiting distributions is proven, with the appropriate rates given above, but explicit analytic forms for the distributions are not available. However, existence of the limiting distributions permits proof of the validity of the bootstrap for making inference and allows derivation of valid asymptotic confidence intervals for $\theta(x, y \mid \Psi)$, $\lambda(x, y \mid \Psi)$ and $\gamma(x, y \mid \Psi)$ in all the cases considered above (that is, using either FDH, VRS-DEA or CRS-DEA estimators as appropriate). Technical details are given in Jeong and Simar (2006), Kneip et al. (2008, 2011) and Simar and Wilson (2011a).

3.3 Inference about Mean Efficiency

For testing hypotheses involving mean efficiency or other model features, more than just the asymptotic results for an individual point (x, y) given above is needed. In most cases, as seen below in Section 4, tests will be based on test statistics involving sample means of efficiency estimators evaluated at each of the random points (X_i, Y_i) contained in the sample χ_n. Hereafter, we drop the notation '$\bullet$' in $\widehat{\theta}_{\bullet,n}$ used above to indicate which estimator is used. That which is derived below is applicable to any estimator (that is, FDH, VRS-DEA or CRS-DEA) and any orientation (that is, input, output or hyperbolic) with the appropriate rate of convergence involving κ as discussed above.

To understand the problem, consider the simplest problem where the researcher wishes to analyze average efficiency over some sector of activity. To do this, we investigate how the sample mean $\widehat{\mu}_n = n^{-1} \sum_{i=1}^{n} \widehat{\theta}_n(X_i, Y_i)$ is related to the population mean $\mu_\theta = \mathbb{E}(\theta(X, Y \mid \Psi)$. It is important to note that the sample mean $\widehat{\mu}_n$ involves *estimators* of efficiency, as opposed to the true efficiencies. The pioneering work of Kneip et al. (2015b) makes clear that for even this seemingly simple problem there is no easy solution due to the inherent bias of the estimators. This bias remains of the same order $n^{-\kappa}$ after taking the mean over n observations, and consequently kills the variance which is proved by Kneip et al. (2015b) to be of order n^{-1}.

The basic, key theorem in Kneip et al. (2015b) may be summarized as follows. Under mild regularity conditions, which differ slightly according the chosen estimator, we have (from Kneip et al., 2015b, Theorem 4.1)

(i) $\widehat{\mu}_n$ is a consistent estimator of μ_θ, with a bias term of order $Cn^{-\kappa}$;
(ii) $\widehat{\sigma}^2_{\theta,n} = n^{-1}\sum_{i=1}^{n}\left[\hat{\theta}_n(X_i, Y_i) - \widehat{\mu}_n\right]^2$ is a consistent estimator of $\sigma^2_\theta = \text{Var}(\theta(X, Y \mid \Psi))$; and
(iii) the following basic limit theorem

$$\sqrt{n}\left(\widehat{\mu}_n - \mu_\theta - Cn^{-\kappa} - R_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2_\theta), \text{ with } R_{n,\kappa} = o(n^{-\kappa}) \tag{4.12}$$

as $n \to \infty$ holds under appropriate regularity conditions.

From these results, it is easy to see why problems arise in making inference about mean efficiency. If $\kappa > 1/2$, there is no problem and a simple Central Limit Theorem (CLT) derived from (4.12) is available, because $\sqrt{n}(Cn^{-\kappa} + R_{n,\kappa}) = o(1)$. However, $\kappa > 1/2$ if and only if $p + q \le 1$ when FDH estimators are used, or $p + q \le 2$ for the VRS-DEA case or $p + q \le 3$ for the CRS-DEA case. If $\kappa = 1/2$ then the bias is a constant, and if $\kappa < 1/2$ the bias explodes. In either of these cases (that is, whenever $\kappa \le 1/2$), standard CLT results (for example, the Lindeberg-Feller CLT) cannot be used. This means that when FDH estimators are used, the usual CLT results are *never* applicable. Moreover, replacing the scale factor $\sqrt{n}$ with n^ξ with $\xi < \kappa \le 1/2$ is not a viable option, since this would cause the variance to converge to zero whenever $\kappa \le 1/2$, rendering inference impossible.

For cases where $\kappa \le 1/2$ (which is typical in applied work), the solution suggested by Kneip et al. (2015b) is twofold. First, one must correct for the leading term $Cn^{-\kappa}$ in the inherent bias in (4.12) using a generalized jackknife estimator of the bias. Second, when necessary, depending on the estimator and on the dimension $p + q$, one must compute the average on a subsample of the n efficiency estimates to control the variance of the resulting average.

The bias correction works along the following lines. As shown in Kneip et al. (2015b), a consistent estimator of the bias term $Cn^{-\kappa}$ is given by

$$\widehat{B}_{n,\kappa} = (2^\kappa - 1)^{-1}(\bar{\mu}^*_{n/2} - \widehat{\mu}_n), \tag{4.13}$$

where $\bar{\mu}^*_{n/2}$ is a jackknife analog of $\widehat{\mu}_n$ obtained as

$$\bar{\mu}^*_{n/2} = (\bar{\mu}^{(1)}_{n/2} + \bar{\mu}^{(2)}_{n/2})/2, \tag{4.14}$$

where $\bar{\mu}^{(1)}_{n/2}$ is a version of $\widehat{\mu}_n$ based on a random subset of χ_n of size $n/2$ and $\bar{\mu}^{(2)}_{n/2}$ is the analog based on the remaining part of the sample (for simplicity, assume that n is even). Formally, one computes for $j = 1, 2$,

$$\bar{\mu}^{(j)}_{n/2} = 2n^{-1} \sum_{\{i \mid (X_i, Y_i) \in \chi^{(j)}_{n/2}\}} \hat{\theta}_{n/2}(X_i, Y_i \mid \chi^{(j)}_{n/2}), \tag{4.15}$$

where $\chi^{(1)}_{n/2} \cap \chi^{(2)}_{n/2} = \varnothing$ and $\chi^{(1)}_{n/2} \cup \chi^{(2)}_{n/2} = \chi_n$. Note carefully that it is explicit in the notation that the efficiency estimators (either FDH, VRS-DEA or CRS-DEA) are computed using the limited reference samples $\chi^{(j)}_{n/2}$. The intuition behind (3.8) is that the $\bar{\mu}^{(j)}_{n/2}$, $j = 1, 2$, are two estimators of μ_θ each having the same asymptotic bias as $\widehat{\mu}_n$, but with n replaced by $n/2$.

Kneip et al. (2015b) prove that $\widehat{B}_{n,\kappa} = Cn^{-\kappa} + o(n^{-\kappa}) + o_p(n^{-1/2})$. Kneip et al. (2016) suggest repeating the exercise just described a large number of times (say, L times) providing L estimates of the bias, and then averaging the L estimates to reduce noise. Then for $\ell = 1, \ldots,$ L, we have, based on the ℓth random share of χ_n in two parts, as above, the estimates $\widehat{B}^{\ell}_{n,\kappa}$ given in (4.13). These are then used in

$$\widehat{B}_{n,\kappa} = L^{-1}\sum_{\ell=1}^{L}\widehat{B}^{\ell}_{n,\kappa} \tag{4.16}$$

to estimate the required bias correction. Averaging over the L bias estimates reduces the variance of the final bias estimate (in practice, selecting $L = 20$ is typically sufficient).

The bias correction solves the main part of the bias problem, but the remainder $R_{n,\kappa} = o(n^{-\kappa})$ must also be controlled. This is achieved, when $p + q$ is too large, by limiting the number of efficiency estimates used to compute $\widehat{\mu}_n$. When $\kappa \leq 1/2$, a subsample version of $\widehat{\mu}_n$ given by

$$\widehat{\mu}_{n_\kappa} = n_\kappa^{-1} \sum_{\{j|(X_j,Y_j)\in \chi^*_{n_\kappa}\}} \widehat{\theta}_n(X_j, Y_j \mid \chi_n), \tag{4.17}$$

must be used, where the average is taken over a random subsample $\chi^*_{n_\kappa} \subseteq \chi_n$ of size $n_\kappa = \lfloor n^{2\kappa} \rfloor \leq n$, where $\lfloor a \rfloor$ denotes the largest integer not greater than $a \in \mathbb{R}$. However, the individual efficiency estimates are computed using the full sample as the reference set, as seen explicitly in the notation. Due to the particular expressions of $R_{n,\kappa}$ which depend on the chosen estimator, Kneip et al. (2015b) derive (under the same regularity assumptions as above) the following CLTs.

(i) For $p + q \leq 3$, if the FDH estimator is used, or for $p + q \leq 4$ and if Ψ is convex and the DEA-VRS estimator is used, or for $p + q \leq 5$ and if Ψ is convex and Ψ^{∂} exhibits CRS and either the CRS-DEA or the VRS-DEA estimator is used, then

$$\sqrt{n}\left(\widehat{\mu}_n - \mu_\theta - \widehat{B}_{n,\kappa} - R_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2_\theta), \text{ with } R_{n,\kappa} = o(n^{-1/2}) \tag{4.18}$$

as $n \to \infty$; and

(ii) when $\kappa < 1/2$, with κ depending on the chosen estimator as discussed above, then

$$\sqrt{n_\kappa}\left(\widehat{\mu}_{n_\kappa} - \mu_\theta - \widehat{B}_{n,\kappa} - R_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2_\theta), \text{ with } R_{n,\kappa} = o(n^{-\kappa}) \tag{4.19}$$

as $n \to \infty$.

Note that either of the two CLTs in (4.18) and (4.19) can be used with the FDH estimator when $p + q = 3$ (or when $p + q = 4$ with VRS-DEA under convexity, or when $p + q = 5$ under convexity and CRS, using either the VRS-DEA or CRS-DEA estimator). But as explained in Kneip et al. (2015b), in these particular cases, the second result in (4.19) provides better approximations in finite samples.

The CLTs (4.12), (4.18) and (4.9) provide tools for building asymptotic confidence intervals for μ_θ. For example, suppose $p + q \geq 4$, Ψ is assumed convex, and VRS-DEA estimators are used. Using (4.19), an asymptotically correct $(1 - \alpha)$ confidence interval for μ_θ is given by

$$\left[\widehat{\mu}_{n_\kappa} - \widehat{B}_{n,\kappa} \pm z_{1-\alpha/2}\widehat{\sigma}_{\theta,n} / n^{\kappa}\right], \tag{4.20}$$

with $\kappa = 2/(p + q + 1)$ and where $z_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard normal distribution. These basic CLTs also provide tools that are useful for solving various testing problems as described below in Sections 4–6. Each test described in Sections 4–6 involves comparison of two sample means of efficiency estimates.

3.4 Dimension Reduction

The curse of dimensionality discussed above in Section 3.2 may jeopardize the quality and the usefulness of both estimates and any statistical inference in nonparametric frontier models if the dimension of the problem, $p + q$, is 'too large' relative to the given sample size n. Wilson (2018) discusses a number of 'rules-of-thumb' for deciding whether dimensionality is 'too large' that have appeared in the literature, and also discusses why these rules are not useful. In many applied problems, substantial multicollinearity exists among the outputs and among the inputs, and this can often be exploited to reduce the dimensionality of the problem. Daraio and Simar (2007) suggest a method for reducing dimensionality based on factor analysis and exploiting near-collinearity among inputs or outputs when it exists. Wilson (2018) provides several diagnostics for assessing whether the number of dimensions in a given problem is excessive for the available sample size, and for indicating whether dimension reduction is likely to reduce estimation error. Wilson (2018) also provides extensive Monte Carlo results to quantify the tradeoff between information lost versus the reduction in estimation error (in terms of mean-square error) due to dimension reduction.

The method developed by Daraio and Simar (2007) is based on the singular value decomposition of a moment matrix. Dimension reduction may be applied to either inputs, outputs or both. Here, we illustrate the idea in terms of inputs. Consider the case where some inputs are closely related. Let $\boldsymbol{X}$ denote the $(n \times k)$ matrix of observations on these inputs, where $k \leq p$. The goal is to reduce dimensionality by transforming these k input variables to a single measure. Suppose the k columns of $\boldsymbol{X}$ have been standardized by dividing each column by its standard deviation. Then the $(k \times k)$ (raw) moment matrix is $\boldsymbol{X'X}$. Let $\xi_1, \ldots, \xi_k$ denote the eigenvalues of $\boldsymbol{X'X}$ in decreasing order, and let $a_1, \ldots, a_k$ denote the corresponding eigenvectors.

All of the data in $\boldsymbol{X}$ lie in the positive orthant $\mathbb{R}_+^k$. It is well known in the statistical literature (for example, see Haerdle and Simar, 2019), that the best linear representation of the data in $\boldsymbol{X}$ in one dimension is given by the first 'input factor' $F_x = \boldsymbol{X}a_1$, and the quality of the representation is measured by its percent of inertia given by $R_x = \xi_1 / \Sigma_{j=1}^k \xi_j$. The latter provides a measure of how close the original data observations in $\boldsymbol{X}$, in the k-dimensional space, are 'near' their projections F_x on the first eigenvector. In other words, the ratio R_x measures the percentage of the information contained in the k columns of $\boldsymbol{X}$ that is shared by the first factor F_x. For instance, if $R_x = 0.95$, then F_x contains 95 percent of the information contained in $\boldsymbol{X}$. So for large values of R_x (which happens when the columns of $\boldsymbol{X}$ exhibit substantial collinearity), one might reasonably replace the k original inputs by this single univariate factor. As noted in Daraio and Simar (2007), the correlations of the original inputs, that have been aggregated, with this input factor indicate how well this new one-dimensional variable represents the original ones. Wilson (2018) uses Monte Carlo experiments to provide guidelines based on R_x and sample size n for determining whether dimension reduction as described here is likely to reduce mean-square error of the efficiency estimates.

Wilson (2018) notes that some (for example, Adler and Golany, 2007) reduce dimensionality by decomposing the correlation matrix for the columns of $\boldsymbol{X}$ instead of the moment matrix

X′X. But for radial efficiency estimated by FDH, VRS-DEA or CRS-DEA estimators, the moment matrix is more appropriate, since it provides information on how tightly packed are rays from the origin to each observation and it is along these rays that efficiency is measured, By contrast, the correlation matrix involves central moments, and describes how the data vary around means of the input (or output) variables, which is irrelevant for purposes of radial efficiency estimates.

4. TESTING HYPOTHESES ABOUT MODEL FEATURES

4.1 Testing Equality of Mean Efficiency for Two Groups of Units

We begin with a relatively easy test. Suppose there are two independent samples of size n_1 and n_2 of firms belonging to two groups labeled G_1 and G_2. We want to test whether $\mu_{1,\theta} = \mathbb{E}_{G_1}(\theta(X,Y) \mid \Psi_1)$ and $\mu_{2,\theta} = \mathbb{E}_{G_2}(\theta(X,Y) \mid \Psi_2)$ are equal against the alternative hypothesis that group 1 has a larger mean efficiency than group 2. Alternatively, one might test whether group 2 has a larger mean efficiency than group 1, or one could simply test whether the two groups have different mean efficiency. The latter amounts to a two-sided test, while the two previous tests are one-sided tests. The test is developed by Kneip et al (2016). Here, we present the most general version of the test, where no restrictions are made on whether the attainable sets for the two groups are the same, that is, whether units in the two groups face the same frontier. Kneip et al. (2016, Section 3.1.2) explain how the test can be adapted to situations where firms in the two groups face the same frontier.

The researcher is confronted with two iid samples, $\chi_{1,n_1} = \{(X_i, Y_i)\}_{i=1}^{n_1}$ and $\chi_{2,n_2} = \{(X_i, Y_i)\}_{i=1}^{n_2}$. Assume the two samples are independent of each other. The CLTs (4.18) or (4.19), depending on the dimension $p + q$ and the chosen estimator (FDH, VRS-DEA or CRS-DEA)) can be applied independently for each of the two groups and the corresponding sample means of estimated efficiencies in each group. For example, if $\kappa < 1/2$, then (4.19) applies and we have for $\ell = 1, 2$, as $n_\ell \to \infty$,

$$\sqrt{n_{\ell,\kappa}}\left(\widehat{\mu}_{n_{\ell,\kappa}} - \mu_{\ell,\theta} - \widehat{B}_{\ell,n_\ell,\kappa} - R_{n_\ell,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2_{\ell,\theta}), \text{ with } R_{n_\ell,\kappa} = o(n_\ell^{-\kappa}), \tag{4.21}$$

where $n_{\ell,\kappa} = \lfloor n_\ell^{2\kappa} \rfloor \le n_\ell$. Necessarily, two sequences of independent random variables each having a normal limiting distribution, possess a joint limiting bivariate normal distribution with zero covariance. Consequently, the difference of the two random independent sequences has a limiting normal distribution.

Using the CLTs in (4.18) and (4.19) developed by Kneip et al. (2015b) leads to test results for two different cases. First, if $p + q \le 3$ when FDH estimators are used, or if $p + q \le 4$, Ψ_1 and Ψ_2 are convex and the VRS-DEA estimator is used, or if $p + q \le 5$, Ψ_1 and Ψ_2 are convex and the frontiers of both sets exhibit CRS and either the CRS-DEA or VRS-DEA estimators are used, then

$$\widehat{\tau}_{1,n_1,n_2} = \frac{(\widehat{\mu}_{1,n_1} - \widehat{\mu}_{2,n_2}) - (\widehat{B}_{1,n_1,\kappa} - \widehat{B}_{2,n_2,\kappa}) - (\mu_{1,\theta} - \mu_{2,\theta})}{\sqrt{\widehat{\sigma}^2_{1,\theta,n_1}/n_1 + \widehat{\sigma}^2_{2,\theta,n_2}/n_2}} \xrightarrow{\mathcal{L}} N(0,1) \tag{4.22}$$

provided $n_1/n_2 \to c$ as $n_1, n_2 \to \infty$ where c is a constant. In the second case, where $\kappa < 1/2$ with κ depending on $p + q$ and the specific estimator used, we have

$$\hat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}} = \frac{(\hat{\mu}_{1,n_{1,\kappa}} - \hat{\mu}_{2,n_{2,\kappa}}) - (\hat{B}_{1,n_1,\kappa} - \hat{B}_{2,n_2,\kappa}) - (\mu_{1,\theta} - \mu_{2,\theta})}{\sqrt{\hat{\sigma}^2_{1,\theta,n_1}/n_{1,\kappa} + \hat{\sigma}^2_{2,\theta,n_2}/n_{2,\kappa}}} \xrightarrow{\mathcal{L}} N(0,1), \qquad (4.23)$$

again provided $n_1/n_2 \to c$, as $n_1, n_2 \to \infty$ where c is a constant and for $\ell = 1, 2$, $n_{\ell,\kappa} = \lfloor n_\ell^{2\kappa} \rfloor \leq n_\ell$ and κ characterizes the rate of convergence in each group and has been defined above.

The results in (4.22) and (4.23) are valid for any values of $\mu_{1,\theta}$ and can be used to build asymptotic confidence intervals for their differences. To test for example H_0: $\mu_{1,\theta} = \mu_{2,\theta}$ against the alternative H_1: $\mu_{1,\theta} > \mu_{2,\theta}$ one can compute either τ_{1,n_1,n_2} or $\tau_{2,n_{1,n_\kappa},n_{2,n_\kappa}}$, as appropriate, while setting $(\mu_{1,\theta} - \mu_{2,\theta}) = 0$ (that is, under the null). Then for a test of size α, H_0 is rejected if the test statistic is larger than $\Phi^{-1}(1 - \alpha)$ where $\Phi^{-1}(\cdot)$ is the standard normal quantile function. Alternatively, the null can be rejected in favor of $H_1' : \mu_{1,\theta} < \mu_{2,\theta}$ when the test statistic is less than $\Phi^{-1}(\alpha)$. For a two-sided test, the null is rejected whenever the test statistic is less than $\Phi^{-1}(\alpha/2)$ or greater than $\Phi^{-1}(1 - \alpha/2)$ for a test of size α. Note that the independence of the two samples χ_{1,n_1} and χ_{2,n_2} is crucial, and avoids complications due to covariances.

Intensive Monte Carlo experiments (see Table 1 in Kneip et al., 2016) indicate the test works reasonably well in practice with finite samples. As expected, the performances of the test (that is, the achieved level and power) increases when the sample sizes increase for all the dimensions $(p + q)$ considered. Of course, for a fixed sample size, the performances decrease when the dimension increases: this advocates in practical problem with real data and suggests use of, if possible, the dimension reduction mentioned above.

4.2 Testing Convexity

Testing convexity is based on comparison of a sample mean of FDH efficiency estimates (in either the input, output or hyperbolic direction) which do not impose convexity, and a sample mean of VRS-DEA efficiency estimates (in the same direction) which impose convexity. The problem here is more complicated than in the preceding case, because we need the two sample means to be independent, and this requires, as suggested by Kneip et al. (2016), a random split of the original sample χ_n into two mutually exclusive, collectively exhaustive subsamples. Kneip et al. (2016) establish CLTs based on this simple splitting device. The procedure described by Simar and Wilson (2020a) allows one to remove the ambiguity resulting from a single random split by repeating the sample splitting a large number of times in order to 'integrate out' the randomness of a single split.

To formalize things, we wish to test the null hypothesis H_1: Ψ is not convex. The assumption of free disposability of inputs and the outputs is maintained. Under the null, both the FDH and the VRS-DEA estimators are consistent, but under the alternative, only the FDH estimator is consistent. The researcher is faced with a single iid sample χ_n. By construction, $\hat{\theta}_{\text{VRS}}(X_i, Y_i \mid \chi_n) \leq \hat{\theta}_{\text{FDH}}(X_i, Y_i \mid \chi_n) \leq 1$, where again we make explicit in the notation the reference sample used to compute the estimates. Therefore $\hat{\mu}_{\text{FDH},n} - \hat{\mu}_{\text{VRS},n} \geq 0$. Under the null we expect $\hat{\mu}_{\text{FDH},n} - \hat{\mu}_{\text{VRS},n}$ to be 'small,' while under the alternative the difference is expected to be 'large.'

Using test statistics based on the difference between two means over all n observations in the sample cannot work, as explained by Kneip et al. (2016). If the null is true then $n^b\left(\widehat{\mu}_{\text{FDH},n} - \widehat{\mu}_{\text{VRS},n}\right)$ converges to a degenerate distribution for any power $b \leq 1/2$ of n. Avoiding this problem requires randomly splitting the original sample into two independent subsamples χ_{1,n_1} and χ_{2,n_2}. The sample mean of the VRS-DEA estimator evaluated at each $(X_i, Y_i) \in \chi_{1,n_1}$ using only observations in χ_{1,n_1} as the reference set is computed, and the sample mean of the FDH estimators evaluated at each $(X_i, Y_i) \in \chi_{2,n_2}$ using only observations in χ_{2,n_1} as the reference set is also computed. Note that here the two efficiency estimators have different rates of convergence, that is, $\kappa_1 = 2/(p + q + 1)$ for the VRS-DEA case and $\kappa_2 = 1/(p + q)$ for the FDH case. Kneip et al. (2016) suggest choosing n_1 and n_2 such that $n_1^{\kappa_1} = \lfloor n_2^{\kappa_2} \rfloor$ and $n_1 + n_2 = 2$.

For the VRS-DEA estimator, we have the analog of the quantities defined in Section 3.3 but with reference sample χ_{1,n_1}. This leads to the sample mean $\widehat{\mu}_{\text{VRS},n_1}$, the sample variance $\widehat{\sigma}^2_{\text{VRS},n_1}$ and the bias correction $\widehat{B}_{\text{VRS},\kappa_1,n_1}$ (for explicit expressions, see Kneip et al., 2016). In situations where $\kappa_1 < 1/2$, the subsampling means

$$\widehat{\mu}_{\text{VRS},n_1,\kappa_1} = n_{1,\kappa_1}^{-1} \sum_{\{(X_i,Y_i) \in \chi^*_{1,n_{1,\kappa_1}}\}} \widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \chi_{1,n_1}) \tag{4.24}$$

are needed, where $n_{1,\kappa_1} = \lfloor n_1^{2\kappa_1} \rfloor < n_1$ for $\kappa_1 < 1/2$ and $\chi^*_{1,n_{1,\kappa_1}}$ is a random subset of χ_{1,n_1} of size n_{1,κ_1}. Note the summation in (4.24) is over the subsample, but the estimator under the summation sign is computed using the full reference set χ_{1,n_1}. Similarly, and without giving all the detailed expressions, using the subsample χ_{2,n_2}, we have for the FDH estimator the corresponding sample mean $\widehat{\mu}_{\text{FDH},n_2}$, sample variance $\widehat{\sigma}^2_{\text{FDH},n_2}$ and the bias correction $\widehat{B}_{\text{FDH},\kappa_2,n_2}$. If $\kappa_2 < 1/2$ the subsample version of the mean given by

$$\widehat{\mu}_{\text{FDH},n_2,\kappa_2} = n_{2,\kappa_2}^{-1} \sum_{\{(X_i,Y_i) \in \chi^*_{2,n_{2,\kappa_2}}\}} \widehat{\theta}_{\text{FDH}}(X_i, Y_i \mid \chi_{2,n_2}) \tag{4.25}$$

is required, where the estimates under the summation sign are computed using all the observations in χ_{2,n_2}, but the summation is over only a subset of these observations, analogous the VRS-DEA case in (4.24).

Depending on the value of $p + q$, different versions of the test statistic are available, since for each mean in the two independent groups, we have the CLTs described in (4.18) and (4.19). For $p + q \leq 3$ we have

$$\widehat{\tau}_{3,n} = \frac{(\widehat{\mu}_{FDH,n_2} - \widehat{\mu}_{VRS,n_1}) - (\widehat{B}_{FDH,n_2,\kappa_2} - \widehat{B}_{VRS,n_1,\kappa_1})}{\sqrt{\widehat{\sigma}^2_{FDH,\theta,n_2}/n_2 + \widehat{\sigma}^2_{VRS,\theta,n_1}/n_1}} \xrightarrow{\mathcal{L}} N(0,1) \tag{4.26}$$

as $n \to \infty$ under the null hypothesis of convexity. For $p + q > 3$, we have

$$\widehat{\tau}_{4,n} = \frac{(\widehat{\mu}_{\text{FDH},n_2,\kappa_2} - \widehat{\mu}_{\text{VRS},n_1,\kappa_1}) - (\widehat{B}_{FDH,n_2,\kappa_2} - \widehat{B}_{VRS,n_1,\kappa_1})}{\sqrt{\widehat{\sigma}^2_{FDH,\theta,n_2}/n_{2,\kappa_2} + \widehat{\sigma}^2_{VRS,\theta,n_1}/n_{1,\kappa_1}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{4.27}$$

as $n \to \infty$ under the null. The null hypothesis of convexity is rejected if the *p*-value $\hat{p}_{\ell} = 1 - \Phi(\hat{\tau}_{\ell,n})$ is suitably small (that is, less than 0.1, 0.05 or 0.01) with $\ell \in \{3, 4\}$ indexing the appropriate statistic (depending on $p + q$) in (4.26) or (4.27).

Splitting the original sample χ_n into two independent subsamples χ_{1,n_1} and χ_{2,n_2} is crucial for obtaining the non-degenerate limiting distributions of the test statistics. Simar and Wilson (2020a) show how a bootstrap can be used to remove the ambiguity resulting from a single random split. The idea is to repeatedly split the sample *s* times, and compute either (i) the sample mean of the appropriate test statistic in (4.26) or (4.27) computed for each of the *s* splits, or (ii) the one-sample Kolmogorov-Smirnov statistic comparing the empirical distribution of the *s* *p*-values corresponding to the test statistics computed for each of the *s* sample splits and the uniform distribution over [0, 1]. Due to the dependence between the *s* sample splits, the distributions of these two test statistics are unknown. However, the bootstrap method provided by Simar and Wilson (2020a) permits easy estimation of *p*-values corresponding to the two statistics obtained from *s* sample-splits.

Results from Monte Carlo experiments provided by Simar and Wilson (2020a) indicate that the procedure works well and achieves better performance than the method based on a single split, in terms of both size and power. The Kolmogorov-Smirnov statistic is also observed to have slightly better properties than the averaged statistic. In practice, for a single sample of real data, the computational burden of the proposed bootstrap is feasible for modern desktop machines with sample sizes typically encountered in applied work.

The test of convexity of Ψ is extended by Kneip et al. (2016) and Simar and Wilson (2020a) for testing CRS against the alternative hypothesis of VRS (under the assumption of convexity for Ψ). The idea is analogous to the idea underlying the convexity test discussed here, and involves comparing sample means of DEA-VRS and DEA-CRS estimators. The issues are also similar; independence between the two sample means is required, which in turn requires randomly splitting the original sample into two independent subsamples. See Kneip et al. (2016) and Simar and Wilson (2020a) for specific details. All of the testing methods discussed so far have been implemented in the current version of the *R* package FEAR introduced by Wilson (2008).

5. DEALING WITH ENVIRONMENTAL FACTORS

5.1 Conditional Frontiers

In applied work, there are often additional variables representing environmental factors that are neither inputs nor outputs in the usual sense, but which may affect production of outputs. When these environmental factors differ across producers, they should be taken into account, as failure to do so may lead to misleading results. Examples of environmental variables faced by local governments might include unobserved heterogeneity across regions within a country, differing legal and regulatory environments across states or provinces, differing forms of government across municipalities (for example, mayors versus city managers) and other factors. These factors are beyond the control of the municipality, but may affect municipalities' ability to transform inputs into outputs in either of two ways. The environmental factors may (i) influence the shape of the attainable frontier Ψ^{∂}, or (ii) influence the probability of being close to the efficient boundary Ψ^{∂} by affecting the shape of the distribution of efficiency or (iii) influence both Ψ^{∂} and the distribution of efficiency.

Let Z denote a random vector of r environmental variables, with $Z \in \mathcal{Z} \subseteq \mathbb{R}^r$. The concept of *conditional frontiers* introduced by Cazals et al. (2002) and extended by Daraio and Simar (2005) provides a natural and appealing way to incorporate information contained in Z.[4] The *conditional* attainable set is

$$\Psi^z = \{(x, y) \mid x \text{ can produce } y \text{ when } Z = z\} \tag{4.28}$$

for given levels $Z = z$ of the environmental factors. The *marginal* attainable set is then

$$\Psi = \{(x, y) \mid x \text{ can produce } y\} = \bigcup_{z \in \mathcal{Z}} \Psi^z. \tag{4.29}$$

In general, $\Psi^z \subseteq \Psi$, which includes the special case

$$\Psi^z = \Psi \text{ for all } z \in \mathcal{Z}. \tag{4.30}$$

This is known as the 'separability' condition as described by Simar and Wilson (2007, 2011b). The separability condition in (4.30) is quite restrictive; it is one case among an uncountable set of possibilities. Simar and Wilson (2007) caution against merely assuming that the separability condition holds, and note that (4.30) should be tested against the alternative hypothesis that $\Psi^z \neq \Psi$ for some $z \in \mathcal{Z}$. However, as explained clearly by Simar and Wilson (2007, 2011b) (4.30) must hold in order for a second-stage regression of efficiency estimates on environmental variables Z to estimate anything sensible and meaningful. If (4.30) does not hold, then such second-stage regressions estimate some unknown thing that is not a feature of any model in which the left-hand side efficiency estimates estimate distance to the frontier of the production set. If (4.30) does not hold, then the FDH, VRS-DEA or CRS-DEA estimators of (4.3)–(4.5) estimate distance to a frontier that may not be attainable and is not relevant.[5]

The separability condition is an important research question whenever environmental variables are available. A formal test of separability versus non-separability is discussed below in Section 5.2. If the condition (4.30) is rejected, the model is said to be 'non-separable,' and in this case conditional efficiency estimators are needed to estimate distance to the boundary of Ψ^z. If the model is non-separable, then the unconditional efficiency estimators discussed above in Section 3 estimate distance to the boundary of Ψ instead of Ψ^z, but $\Psi^z \subset \Psi$ for some $z \in \mathcal{Z}$. In addition, conditional efficiency estimators are also needed to build the test of separability below in Section 5.2.

Conditional efficiency measures are defined along the lines of the usual 'marginal' scores defined above in (4.3)–(4.5). Here, in the conditional case, the attainable set is restricted to Ψ^z. For example, for the input orientation, we have the conditional efficiency measure

$$\theta_c(x, y \mid z) = \theta(x, y \mid \Psi^z) = \inf\{\theta > 0 \mid (\theta x, y) \in \Psi^z\}. \tag{4.31}$$

Since by (4.29), $\forall$ z, $\Psi^z \subseteq \Psi$, we have $\theta(x, y) \leq \theta_c(x, y \mid z) \leq 1$.

Nonparametric estimators are based on an iid sample $\mathcal{S}_n = \{(X_i, Y_i, Z_i)\}_{i=1}^n$ of observations on input and output quantities and corresponding environmental variables. The usual, unconditional FDH and DEA estimators are computed by estimating Ψ, the marginal support of (*X*, *Y*), by the appropriate envelopment of the marginal cloud of points $\chi_n = \{(X_i, Y_i)\}_{i=1}^n$ as in (4.19)–(4.21). Here, it is important to make the reference sample used for estimation explicit in the notation. Hence, we write, as above for the marginal (input oriented) estimators,

$$\hat{\theta}(X_i, Y_i) = \hat{\theta}(X_i, Y_i \mid \chi_n) \tag{4.32}$$

and similarly for the output and hyperbolic orientations.

The conditional efficiency estimators require FDH (or VRS-DEA or CRS-DEA) estimators of Ψ^z. The easiest case is when the variables in *Z* are discrete, with several firms observed for each value of the discrete variable. In this case, for any possible value *z* of *Z*, we have a sample of observations (*X*, *Y*) corresponding to the cases where $Z = z$. This defines a subsample $\chi_n^z = \{(X_{i_j}, Y_{i_j}) \mid Z_{i_j} = z\}_{j=1}^{n_z} \subset \chi_n$ of size $n_z < n$, where $n_z = \sum_{i=1}^n \mathbb{I}(Z_i = z)$ and $\mathbb{I}(\cdot)$ denotes the indicator function. This cloud of 'conditional' data points allows estimation of Ψ^z, providing FDH, VRS-DEA or CRS-DEA versions of the conditional efficiency scores by adapting (4.19)–(4.21) to this subsample. Hence, we denote the conditional (input-oriented) estimator of the efficiency of an observation (X_i, Y_i, Z_i) as

$$\hat{\theta}_c(X_i, Y_i \mid Z_i) = \hat{\theta}(X_i, Y_i \mid \chi_n^{Z_i}). \tag{4.33}$$

Conditional efficiency estimators have also been developed for the case of continuous *Z*. In the continuous case, smoothing techniques are needed to 'localize' the FDH and DEA estimators so that only observations (X_i, Y_i) having corresponding values of Z_i lying in an appropriate neighborhood of *z* are used to estimate Ψ^z. For example, we compute the conditional FDH estimator of the unit facing the environmental conditions $Z = z$ by considering only the free disposal hull (or the free disposal and convex hull for DEA) of the data $\chi_{n,h}^z = \{(X_i, Y_i) \mid z - h \le Z_i \le z + h\}$, where the vector *h* of 'bandwidths' tunes (or controls) the neighborhood of *z*. The VRS-DEA and CRS-DEA estimators are computed similarly, using only the convex or conical hulls of the free disposal hull of observations (X_i, Y_i) such that the corresponding observations Z_i satisfy $z - h \le Z_i \le z + h$. The bandwidths must be small enough to describe units facing environmental conditions similar to *z* (thereby minimizing some potential bias), but sufficiently large so as to have enough observations in $\chi_{n,h}^z$ to avoid excessive variance. As in many nonparametric problems, bandwidths can be chosen to optimize mean integrated square error, which reflects the tradeoff between bias and variance. As the bandwidths are made smaller (larger), bias is reduced (increased) while variance is increased (reduced). Theoretical results given by Jeong et al. (2010), Daraio et al. (2018), Bădin et al. (2019) and references cited therein establish that the *j*th optimal bandwidth, corresponding to the *j*th element of *Z*, $j = 1, \ldots, r$ must be of order $h_j = c_j n^{-1/(r+4)}$. This leads to conditional efficiency estimators with rate of convergence $(nh^r)^\kappa = n^{4\kappa/(r+4)} < n^\kappa$ where $\kappa = 1/(p + q)$, $2/(p + q + 1)$ or $2/(p + q)$) for conditional FDH, VRS-DEA or CRS-DEA estimators, respectively. For technical details and asymptotic properties, see Jeong et al. (2010), Daraio et al. (2018) and Bădin et al. (2019). Daraio et al. (2018) provide CLTs for means of conditional efficiency scores that are used to derive the separability test for continuous *Z*.

5.2 Testing Separability

Simar and Wilson (2007) noted that the separability condition in (4.30) is very restrictive and should be tested, but a development of a formal test had to wait for the additional results of Kneip et al. (2015b). Daraio et al. (2018) developed CLTs for conditional efficiency measures, and used these to develop a test of separability for general cases where Z is either continuous or discrete.[6] To simplify we explain here the basic ideas of the test in the presence of a single discrete variable Z. To establish necessary notation, let $(z_1, \ldots, z_K)$ be the set of possible values for the variable Z. Then the random sample $\mathcal{S}_n$ provides n_k units with $Z_i = z_k$ for $k = 1, \ldots, K$ with $n_1 + \ldots + n_K = n$. The applied researcher has the marginal cloud of points in χ_n as well as the conditional clouds of points $\chi_n^k = \{(X_{i_j}, Y_{i_j}) \mid Z_{i_j} = z_k\}_{j=1}^{n_k}$ for $k = 1, \ldots, K$.

The idea is to build a test statistic based on the difference between $\mu_\theta = \mathbb{E}(\theta(X,Y))$, the mean of the efficiency defined relative to the support of the marginal set Ψ, and $\mu_{\theta,c} = \mathbb{E}(\theta_c(X,Y \mid Z))$, the mean of the conditional efficiency defined relative to the support of the conditional sets $\Psi^{z_k}, k = 1, \ldots, K$. Under the null hypothesis $H_0 : \Psi^{z_k} = \Psi, \forall\, k = 1, \ldots, K$, the difference $\mu_\theta - \mu_{\theta,c}$ is zero. In addition, as noted by Daraio et al. (2018), we know that $\mu_{\theta,c} \geq \mu_\theta$, and hence a test statistic can be derived from the difference $\hat{\mu}_{n,\theta,c} - \hat{\mu}_{n,\theta}$ between the corresponding sample means. These sample means require appropriate estimators of the two population means, where, by construction, $\hat{\mu}_{n,\theta,c} - \hat{\mu}_{n,\theta} \geq 0$. The null hypothesis of separability should be rejected if this difference is 'too big.'

Estimates of individual unconditional efficiencies are given by $\hat{\theta}(X_i Y_i \mid \chi_n)$, $i = 1, \ldots, n$, and for the conditional case we have by (4.33)

$$\hat{\theta}_c(X_i Y_i \mid Z_i) = \hat{\theta}(X_i Y_i \mid \chi_n^k) \tag{4.34}$$

for $i = 1, \ldots, n$ and k such that $Z_i = z_k$. Here either FDH, VRS-DEA or CRS-DEA estimates can be used as appropriate. Similar expressions exist for other orientations (for example, output or hyperbolic). However, a test statistic constructed from sample means of the two sets of estimates over $i = 1, \ldots, n$ cannot be used, since under the null, the asymptotic distribution of $\hat{\mu}_{n,\theta,c} - \hat{\mu}_{n,\theta}$ will be degenerate with mass one at zero as shown by Daraio et al. (2018).

The solution to this problem is analogous to the method used for testing the convexity of Ψ described in Section 4.2. We can randomly split the sample $\mathcal{S}_n$ into two parts $\mathcal{S}_{1,n_1}$ and $\mathcal{S}_{2,n_2}$ such that $n_1 = \lfloor n/2 \rfloor$, $n_2 = n - n_1$ such that $\mathcal{S}_{1,n_1} \cup \mathcal{S}_{2,n_2} = \mathcal{S}_n$ and $\mathcal{S}_{1,n_1} \cap \mathcal{S}_{2,n_2} = \varnothing$.[7] The n_1 observations in $\mathcal{S}_{1,n_1}$ are used to provide the unconditional estimates, while the n_2 observations in $\mathcal{S}_{2,n_2}$ are used to provide the conditional estimators. The sample $\mathcal{S}_{1,n_1}$ provides the marginal cloud of points χ_{1,n_1} and the sample $\mathcal{S}_{2,n_2}$ defines the K conditional clouds of points $\chi_{2,n_2}^k = \{(X_{i_j}, Y_{i_j}) \mid Z_{i_j} = z_k\}_{j=1}^{n_{2,k}}$ for $k = 1, \ldots, K$, where $(X_{i_j}, Y_{i_j}, Z_{i_j}) \in \mathcal{S}_{2,n_2}$.

Estimators of the unconditional and conditional means are given by

$$\hat{\mu}_{n_1} = n_1^{-1} \sum_{(X_i, Y_i, Z_i) \in \mathcal{S}_{1,n_1}} \hat{\theta}(X_i Y_i \mid \chi_{1,n_1}) \tag{4.35}$$

and

$$\hat{\mu}_{c,n_2} = n_2^{-1} \sum_{(X_i, Y_i, Z_i) \in \mathcal{S}_{2,n_2}} \hat{\theta}(X_i Y_i \mid \chi_{2,n_2}^{Z_i}). \tag{4.36}$$

As noted by Daraio et al. (2018), the conditional estimates in group 2 have the usual convergence rate of the unconditional efficiency estimates in group 1. This is because for discrete Z, no bandwidth is involved in the estimation, and each subsample χ^k_{2,n_2} has a size proportional to n_2, for $k = 1, \ldots, K$. Splitting the sample ensures that the two sample means are independent, avoiding difficulties that would otherwise arise from any non-zero covariance.

Given the independent sample means in (4.35) and (4.36), a statistic for testing separability can be constructed along the lines of the difference-in-means test described above in Section 4.1 by adapting the test statistics defined in (4.22) and (4.23). The bias correction terms for group 2 must be computed separately and independently for each subgroup $k = 1, \ldots, K$. This involves splitting each of the K subgroups of group 2 to compute the generalized jackknife estimates of bias for each subgroup as described above. See Daraio et al. (2018) for additional details and proofs. Conditional efficiency estimators, as well as the separability tests of Daraio et al. (2018), are implemented in the FEAR software described by Wilson (2008).

6. DEALING WITH DYNAMIC ISSUES

Malmquist productivity indices are widely used in dynamic contexts for measurement of productivity change from one period to another. They are often estimated using nonparametric, data envelopment analysis (DEA) estimators due to the work of Färe et al. (1992). Kneip et al. (2021) provide the theoretical developments needed to make inference about productivity change measured by Malmquist indices when Ψ is convex, for both individual producers as well as for geometric means over a sample of producers. Malmquist indices are often decomposed into various measures of the sources of productivity change, such as changes in efficiency or changes in technology (that is, movement of the efficient frontier over time). Simar and Wilson (2019) provide CLTs that permit inference about geometric means of the components of Malmquist indices, as well as arithmetic means of their logarithms. We summarize here the results providing asymptotic confidence intervals for changes in efficiency and for changes in technology that are used in our application below in Section 7. For more details, see Simar and Wilson (2019) and Kneip et al. (2021).

Assume that at time t, the attainable set is given by

$$\Psi^t = \{(x, y) \mid x \text{ can produce } y \text{ at time } t\}. \tag{4.37}$$

Then the efficient frontier at time t is given by

$$\Psi^{t\partial} = \{(x, y) \in \Psi^t \mid (\gamma^{-1}x, \gamma y) \notin \Psi^t \ \forall\, \gamma > 1\}. \tag{4.38}$$

As in Section 2, we assume the same, usual characteristics for Ψ^t coming from production theory.

Now consider a sample of inputs and outputs for n units observed in periods $t = 1$ and $t = 2$. Denoting $W_i^t = (X_i^t, Y_i^t)$, the sample is represented by $\chi_n = \{(W_i^1, W_i^2)\}_{i=1}^n$. Note that the W_i^t are independent for $i = 1, \ldots, n$, but may be dependent across time, that is, W_i^1 and W_i^2 may be dependent. For notational simplicity, assume here and below that the same units are observed in both time periods (Simar and Wilson, 2019 consider the more general situation where some units are observed in only one period). We analyze the components of the Malmquist indices described in equation (2.11) in Simar and Wilson (2019), in terms of hyperbolic measures of efficiency defined in (4.5). Working in the hyperbolic orientation avoids problems that may arise when working in the input or output directions, where some cross-period efficiencies might not be defined. The change in efficiency for the ith unit from t_1 to t_2 is measured by

$$\mathcal{E}_i = \frac{\gamma(W_i^2 \mid \Psi^2)}{\gamma(W_i^1 \mid \Psi^1)}, \tag{4.39}$$

and the technology change from t_1 to t_2 faced by the ith unit is described by

$$\mathcal{T}_i = \left(\frac{\gamma(W_i^2 \mid \Psi^1)}{\gamma(W_i^2 \mid \Psi^2)} \times \frac{\gamma(W_i^1 \mid \Psi^1)}{\gamma(W_i^1 \mid \Psi^2)} \right)^{1/2}. \tag{4.40}$$

Note that both $\mathcal{E}_i$ and $\mathcal{T}_i$ provide measures of change regardless of whether the attainable sets are convex or whether returns to scale are constant or otherwise. Consequently, estimators $\widehat{\mathcal{E}}_i$ and $\widehat{\mathcal{T}}_i$ of $\mathcal{E}_i$ and $\mathcal{T}_i$ can be obtained by replacing the hyperbolic distance measures on the right-hand sides of (4.39) and (4.40) with FDH, VRS-DEA or CRS-DEA estimators, depending what is assumed about Ψ^t for $t \in \{1, 2\}$. Then the convergence rates for the resulting estimators of $\mathcal{E}_i$ and $\mathcal{T}_i$ are the same as the rate for the particular estimator substituted on the right-hand side.

Simar and Wilson (2019) extend the results of Kneip et al. (2021) to derive the asymptotic behavior of these estimators which share the same rate of convergence n^κ of the efficiency estimators used to construct $\widehat{\mathcal{E}}_i$ and $\widehat{\mathcal{T}}_i$, where corresponding values of κ are given in Section 3 and depend on the particular estimator used (that is, FDH, VRS-DEA or CRS-DEA). These results permit estimation of asymptotic confidence intervals for $\mathcal{E}_i$ and $\mathcal{T}_i$ using the subsampling methods described in Simar and Wilson (2011a).

Researchers often report geometric means of estimates of changes in efficiency or technology over all of the n units represented in a given sample. For efficiency change, applied researchers may compute the geometric mean

$$\widehat{G}_{\mathcal{E},n} = \left[\prod_{i=1}^{n} \widehat{\mathcal{E}}_i \right]^{1/n} = \exp\left(n^{-1} \sum_{i=1}^{n} \log(\widehat{\mathcal{E}}_i) \right), \tag{4.41}$$

which can be viewed as an estimator of $G_{\mathcal{E}} = \exp\left(\mathbb{E}[\log(\mathcal{E}_i)]\right)$. The geometric mean of technical change estimates $\widehat{\mathcal{T}}_i$, $i = 1, \ldots, n$ can be computed similarly. Simar and Wilson (2019, Theorem 4.10) derive limit theorems for these geometric means and other variants. These CLTs are analogous, *mutatis mutandis*, to the CLTs presented in (4.22) and (4.23), with limiting normal distributions with the rate of convergence governed by κ (and with the value of κ varying according the chosen nonparametric estimator of the individual efficiencies and the dimension $p + q$ of the production set).

For example, for the geometric mean of the estimated efficiency changes, we have

$$\sqrt{n}\left(\widehat{G}_{\mathcal{E},n} - G_{\mathcal{E}} - \widehat{B}_{\mathcal{E},n,\kappa} - R_{\mathcal{E},n,\kappa} \right) \xrightarrow{\mathcal{L}} N(0, G_{\mathcal{E}}^2 \sigma_{\mathcal{E}}^2), \text{ with } R_{\mathcal{E},n,\kappa} = o(n^{-1/2}) \tag{4.42}$$

as $n \to \infty$ provided: (i) $p + q \leq 3$ if the FDH estimator is used; (ii) $p + q \leq 4$ and Ψ^t is convex if the VRS-DEA estimator is used; or (iii) $p + q \leq 5$ if Ψ^t is convex, CRS holds, and either the VRS-DEA or CRS-DEA estimator is used. Alternatively, if $\kappa < 1/2$ then

$$\sqrt{n^{\kappa}}\left(\widehat{G}_{\mathcal{E},n_{\kappa}} - G_{\mathcal{E}} - \widehat{B}_{\mathcal{E},n,\kappa} - R_{\mathcal{E},n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, G_{\mathcal{E}}^2 \sigma_{\mathcal{E}}^2), \text{with } R_{\mathcal{E},n,\kappa} = o(n^{-\kappa}) \tag{4.43}$$

as $n \to \infty$ under appropriate assumptions on Ψ^t corresponding to the particular estimator used (for example, FDH, VRS-DEA or CRS-DEA).

The bias correction $\widehat{B}_{\mathcal{E},n,\kappa}$ is computed using a generalized jackknife procedure similar to the one described in Section 3. The sample χ_n must be randomly split into two subsamples as before to ensure independence, but here the split is done across the set of n firms, keeping the same pair of observations for a given firm at times t_1 and t_2 in the same subsample to preserve the time-dependence (see Simar and Wilson, 2019 for details). Since we have consistent estimators of $G_{\mathcal{E}}$ and of $\sigma_{\mathcal{E}}^2$, the limit theorems in (4.42) and (4.43) can be used to construct asymptotic confidence intervals for $G_{\mathcal{E}}$ by using quantiles of the normal distribution. Equivalently, as is standard in statistical practice, one can test the null hypothesis that $G_{\mathcal{E}} = 1$ (i.e, no efficiency changes between the two periods) versus the alternative hypothesis that $G_{\mathcal{E}} \neq 1$, by checking whether the appropriate estimated confidence interval includes the value 1. Estimates of the measures of efficiency and technology change in (4.39) and (4.40), as well as the Malmquist index and its other components, can be computed using the FEAR library introduced by Wilson (2008).

7. EFFICIENCY OF US MUNICIPAL GOVERNMENTS

In this section we show how the tools discussed above can be used to analyze the performance of US municipal governments. We examine performance in two years, 2002 and 2012. Data are taken from the Annual Survey of State and Local Government Finances, the Annual U.S. Building and Permit Survey, the U.S. Census of Governments, the U.S. Bureau of Labor Statistics (BLS) and the Federal Bureau of Investigation (FBI) to define input and output variables broadly representative of those used in studies of local governments' efficiency in the literature. All dollar amounts are measured in terms of thousands of constant, 2008 US dollars. We use the U.S. Consumer Price Index (CPI) reported by the BLS for cities and regions to adjust for price differences across municipalities in the United States.[8]

Studies of local governments are often limited by data availability, and the story here is no different. Nonetheless, we use total current operating expenditures as our single input (denoted by X) so that $p = 1$. This is consistent with other studies, for example, Šťastná and Gregor (2015). We define $q = 6$ output variables consisting of total population (Y_1), total charges for sewerage and waste management (Y_2), the reciprocal of the total crime rate (Y_3), total land area (in square miles) (Y_4), total building permits (Y_5) and the employment rate (Y_6). These are again representative of what others have specified. The broad collection of outputs reflects the wide variety of goods and services provided by municipal governments in the US. Total population reflects the number of people served, and is widely used in studies of local governments (for example, lo Storto, 2013). The reciprocal of the total crime rate reflects law enforcement services providing public safety. Total land area gives a measure of the physical space over which goods and services are provided. Total building permits provides a measure of administrative services provided, and the employment rate gives a measure of amenities (including jobs, but this is also related to other amenities that attract employers). We use the U.S. Census Bureau's definition of local government, corresponding to type code 2 in the U.S. Census of Governments' 14-digit government ID code.

After eliminating observations with missing values for one or more of our variables, we have 730 observations for 2002 and 800 observations for 2012, giving a total of 1530 observations. As one might expect, there is a good deal of collinearity among the six outputs. We standardize the outputs by dividing each variable by its standard deviation, and then compute the eigenvalues of the (raw) moment matrix for the six standardized outputs as discussed in Section 3.4. The ratio of the largest eigenvalue to the sum of the six eigenvalues is 0.9851, providing clear evidence in view of Wilson (2018) that dimension-reduction is likely to reduce the mean-square error of our efficiency estimates. Consequently, we multiply the $(n \times 6)$ matrix of standardized outputs by the eigenvector corresponding to the largest eigenvalue to construct a single output measure denoted by Y^* as discussed above in Section 3.4. In all of our estimation, we use the single input measure X and the single output measure Y^*, and hence $p = q = 1$. Summary statistics for these variables as well as the original six output measures are provided in Table 4.1. The summary statistics reveal that several of the marginal distributions are heavily skewed to the right, reflecting the distribution of city sizes in the Unites States. Table 4.2 gives the number of observations for each year by census region; regions 1–4 correspond to the Northeastern, Midwestern, Southern and Western parts of the United States.[9]

Table 4.1 Summary statistics

Variable	Min	Q1	Median	Mean	Q3	Max
X	9.4800(2)	5.7970(4)	1.0492(5)	2.6124(5)	2.2749(5)	1.0844(7)
Y_1	7.9700(2)	2.9383(4)	4.5486(4)	8.6136(4)	7.7029(4)	3.7926(6)
Y_2	1.0381(1)	7.6186(3)	1.4982(4)	3.1267(4)	2.8571(4)	1.0198(6)
Y_3	2.8500(−6)	1.7451(−4)	2.5331(−4)	3.2229(−4)	3.7059(−4)	9.5493(−3)
Y_4	5.4000(−1)	1.1950(1)	2.1440(1)	4.2049(1)	3.8172(1)	2.7167(3)
Y_5	1.0000(0)	4.9000(1)	1.7300(2)	4.1293(2)	4.4800(2)	8.8460(3)
Y_6	8.0800(1)	9.1900(1)	9.3600(1)	9.3235(1)	9.5000(1)	9.8500(1)
Y^*	8.1739(2)	3.0965(4)	4.8038(4)	9.1560(4)	8.0390(4)	3.9234(6)

Note: Values are given in scientific notation; for example, 8.1739(2) represents $8.1739 \times 10^2 = 817.39$.

Table 4.2 Sample sizes by year and U.S. Census region

Year	All Regions	Region 1	Region 2	Region 3	Region 4
2002	730	75	158	289	208
2012	800	72	200	299	229
TOTAL	1530	147	358	588	437

We first estimate efficiency for each year across all regions, and then by region using only observations in a specific year and in region $j \in \{1, 2, 3, 4\}$ to estimate hyperbolic efficiency given by (4.5) for observations in the same year and region. We use first the original input-output specification with X and $Y_1, \ldots, Y_6$ and then with the reduced-dimension data consisting of X and Y^*, and we consider the FDH, VRS-DEA and CRS-DEA estimators. We report in Table 4.3 the number of observations with efficiency estimates equal to one in each case. Using the FDH estimator on the original data, almost half of the observations in each year yield estimates equal to one when data are pooled across regions. Estimating independently for each region, the FDH estimator yields estimates equal to one for well over half of the observations in a region-year. The VRS-DEA estimator yields far fewer estimates equal to one, and the CRS-VRS estimator yields still fewer such estimates. This is to be expected, but it is important to realize that if one used only the VRS-DEA estimator, most of the apparent inefficiency one would find is only a consequence of the VRS-DEA estimator's imposition of convexity on Ψ. But whether Ψ is convex is an empirical question.

Before testing convexity of Ψ, we first test separability with respect to census regions for each year, and then we test separability with respect to time for each census region. For these tests, we use the reduced-dimension data and input, output and hyperbolic oriented FDH estimators. We use the FDH estimators since they require neither convexity nor constant returns to scale, and thus are the most flexible among the three types of nonparametric estimators. We use 100 sample-splits and 1000 bootstrap replications as described in Section 5.2. Results are given in Tables 4.4 and 4.5, where 'Test #1' indicates tests based on averaging test statistics across the sample splits, and 'Test #2' indicates tests based on the distribution of the p-values for each split as discussed in Section 5.2. For the tests of separability with respect to census regions we have six tests, and each yields a p-value of zero as shown in Table 4.4. For the tests of separability with respect to time by census regions, we have 24 tests, and again all of the 24 p-values are equal to zero. Consequently, the tests provide strong evidence against separability

Table 4.3 Numbers of observations with estimated hyperbolic technical efficiency equal to 1 in each year

			Without Dimension — Reduction —			With Dimension — Reduction —		
Year	Region	n	FDH	VRS	CRS	FDH	VRS	CRS
2002	all	730	364	35	6	28	5	1
2012	all	800	381	35	7	31	6	1
2002	1	75	61	17	5	9	3	1
	2	158	99	23	4	23	6	1
	3	289	186	44	6	27	7	1
	4	208	148	27	8	19	5	1
2012	1	72	64	23	6	10	3	1
	2	200	118	27	4	23	9	1
	3	299	186	42	5	26	5	1
	4	229	167	33	13	25	5	1

Table 4.4 Test for separability with respect to census regions, by time (100 splits, 1000 bootstrap replications; p = q = 1)

	— Input —		— Output —		— Hyperbolic —	
Year	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value
Test #1:						
2002	9.301	0.000	7.157	0.000	7.806	0.000
2012	11.021	0.000	9.131	0.000	11.814	0.000
Test #2:						
2002	0.919	0.000	0.947	0.000	0.928	0.000
2012	0.997	0.000	0.987	0.000	0.968	0.000

with respect to either time or region. The evidence suggests that in a given year, municipalities in different regions face different frontiers, and in a given region, these frontiers are changing over time (this will be revisited below).

The results of the separability tests make clear that convexity should be tested for each region-year. Table 4.6 gives the results of these tests. Again, we have two tests, #1 and #2, as mentioned above, and we again use 100 sample splits and 1000 bootstrap replications. Altogether, there are 48 tests, and the results are mixed. For example, for region 4 in 2012, convexity is clearly rejected in five out of six cases; the only case in which convexity is not rejected is Test #1 using the output orientation. We test in all three directions as a check on robustness, and the results illustrate that conflicting results can arise. The failure to reject in the output orientation may be a consequence of the heavy right-skewness of X and Y^*, but in any case, there is substantial evidence against convexity for this region-year. By contrast, we

Table 4.5 Test for separability with respect to time by census regions (100 splits, 1000 bootstrap replications; p = q = 1)

	— Input —		— Output —		— Hyperbolic —	
Region	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value
Test #1:						
1	0.182	0.000	5.389	0.000	3.842	0.000
2	0.805	0.000	2.466	0.000	1.636	0.000
3	1.035	0.000	6.015	0.000	6.779	0.000
4	9.527	0.000	15.129	0.000	22.608	0.000
Test #2:						
1	0.333	0.000	0.731	0.000	0.608	0.000
2	0.382	0.000	0.564	0.000	0.478	0.000
3	0.443	0.000	1.000	0.000	1.000	0.000
4	1.000	0.000	0.986	0.000	0.984	0.000

Table 4.6 Test for convexity by year and census region (100 splits, 1000 bootstrap replications; p = q = 1)

		— Input —		— Output —		— Hyperbolic —	
Year	Region	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value
Test #1:							
2002	1	0.415	0.768	1.307	0.995	0.212	0.220
	2	0.712	0.752	0.683	0.306	1.234	0.931
	3	0.012	0.555	0.040	0.666	0.171	0.544
	4	1.896	0.000	0.678	0.966	1.801	0.000
2012	1	0.903	0.916	0.022	0.796	1.549	0.998
	2	0.637	0.209	1.139	0.030	0.513	0.772
	3	1.067	0.001	0.274	0.577	0.242	0.041
	4	1.177	0.000	1.966	1.000	1.961	0.000
Test #2:							
2002	1	0.273	0.156	0.374	0.109	0.233	0.347
	2	0.318	0.257	0.340	0.293	0.383	0.141
	3	0.163	0.695	0.187	0.534	0.139	0.829
	4	0.559	0.001	0.318	0.076	0.537	0.000
2012	1	0.359	0.063	0.266	0.445	0.457	0.003
	2	0.513	0.009	0.477	0.019	0.326	0.247
	3	0.504	0.001	0.228	0.509	0.245	0.504
	4	0.552	0.000	0.594	0.000	0.707	0.000

fail to reject convexity for region 1 in 2002. Since we reject convexity in some, although not all, cases we subsequently use only FDH estimators, as they remain consistent in either case, that is, regardless of whether Ψ is convex, whereas the VRS-DEA and CRS-DEA estimators do not. Moreover, the results of the separability tests discussed above indicate that efficiency should be estimated independently for each region-year.

In Tables 4.7–4.9 we give summary statistics for hyperbolic, input and output oriented FDH efficiency estimates in each region-year. These results reveal a wide range of performances by municipalities in each region-year. Comparison across orientations also reveals some differences and, in particular, the hyperbolic measure suggests greater efficiency on average than either the input or output measures. The hyperbolic measure defined in (4.5) is less sensitive to the curvature of the frontier where an inefficient observation is projected onto the frontier than either the input or the output measures for reasons discussed by Wilson (2011). It is worth noting that our input measure X is total operating expenditure, that is, cost. Consequently, the input-oriented estimates in Table 4.8 can be viewed as estimates of cost-efficiency as discussed

Table 4.7 Summary statistics for hyperbolic FDH technical efficiency estimates by year and census region (with dimension reduction; p = q = 1)

Year	Min	Q1	Median	Mean	Q3	Max	Var
— Region 1: Northeast —							
2002	0.454	0.566	0.669	0.705	0.884	1.000	0.033
2012	0.407	0.650	0.741	0.764	0.916	1.000	0.028
— Region 2: Midwest —							
2002	0.488	0.729	0.825	0.825	0.949	1.000	0.018
2012	0.370	0.722	0.800	0.797	0.891	1.000	0.018
— Region 3: South —							
2002	0.272	0.604	0.742	0.728	0.860	1.000	0.031
2012	0.275	0.617	0.767	0.747	0.890	1.000	0.029
— Region 4: West —							
2002	0.278	0.561	0.677	0.694	0.837	1.000	0.031
2012	0.301	0.539	0.696	0.701	0.855	1.000	0.038

by Simar and Wilson (2020b). While this is true, it is also worth considering other paths to the frontier of the set of feasible costs and output quantities as given in Tables 4.7 and 4.9.

NOTE: Reported values reflect inverse efficiency estimates in order to facilitate comparison with hyperbolic and input-oriented efficiency estimates.

Table 4.8 Summary statistics for input-oriented FDH technical efficiency estimates by year and census region (with dimension reduction; p = q = 1)

Year	Min	Q1	Median	Mean	Q3	Max	Var
— Region 1: Northeast —							
2002	0.190	0.297	0.415	0.525	0.739	1.000	0.072
2012	0.195	0.369	0.590	0.611	0.829	1.000	0.068
— Region 2: Midwest —							
2002	0.166	0.477	0.640	0.667	0.897	1.000	0.053
2012	0.104	0.440	0.584	0.609	0.773	1.000	0.052
— Region 3: South —							
2002	0.046	0.302	0.513	0.530	0.737	1.000	0.073
2012	0.049	0.311	0.515	0.545	0.761	1.000	0.072
— Region 4: West —							
2002	0.028	0.201	0.349	0.439	0.637	1.000	0.081
2012	0.039	0.242	0.361	0.472	0.731	1.000	0.086

Table 4.9 Summary statistics for output-oriented FDH technical efficiency estimates by year and census region (with dimension reduction; p = q = 1)

Year	Min	Q1	Median	Mean	Q3	Max	Var
— Region 1: Northeast —							
2002	0.209	0.368	0.481	0.559	0.750	1.000	0.063
2012	0.153	0.352	0.571	0.604	0.876	1.000	0.080
— Region 2: Midwest —							
2002	0.248	0.571	0.725	0.718	0.893	1.000	0.040
2012	0.174	0.557	0.688	0.688	0.814	1.000	0.039
— Region 3: South —							
2002	0.105	0.410	0.593	0.597	0.785	1.000	0.058
2012	0.127	0.424	0.623	0.619	0.815	1.000	0.057
— Region 4: West —							
2002	0.101	0.435	0.547	0.593	0.735	1.000	0.050
2012	0.112	0.402	0.565	0.598	0.776	1.000	0.055

Tables 4.10 and 4.11 report geometric means of estimates of the efficiency-change and technology-change measures defined in (4.39) and (4.40), along with corresponding *p*-values indicating significant differences from 1 as discussed above in Section 6. Once again, the input, output and hyperbolic directions are used as a robustness check. In Table 4.10, the results give clear evidence of efficiency change in region 1. For the input and hyperbolic orientations, values greater than 1 indicate an increase in inefficiency, while for the output orientation, values less than 1 indicate an increase in inefficiency. Each of the three estimates for region 1 indicate a significant increase in efficiency from 2002 to 2012. Tests using the output orientation are insignificant for regions 2–4, and for region 3 using the hyperbolic orientation. Results using the input orientation are significant for each region, and using the hyperbolic orientation leads to significant results not just for region 1 but also for regions 2 and 4. For region 2, the input and hyperbolic directions agree and suggest a decline in efficiency from 2002 to 2012. The results for regions 3 and 4 are mixed, and so one should be careful about drawing conclusions for these regions as the evidence is not clear.

Table 4.10 Estimates of efficiency change, 2002–12, by region (FDH; p = q = 1)

	— Input —		— Output —		— Hyperbolic —	
Region	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value
1	1.121	9.668(−7)	0.991	6.389(−2)	1.065	6.564(−12)
2	0.968	5.700(−6)	1.009	1.954(−1)	0.991	4.226(−2)
3	0.984	7.298(−3)	0.997	6.075(−1)	1.001	3.184(−1)
4	1.048	2.953(−5)	1.036	9.609(−1)	0.970	2.480(−3)

Note: Values in parentheses give value of exponent in scientific notation; for example, 9.668×10^{-7} is denoted by 9.668(−7).

Table 4.11 Estimates of technology change, 2002–12, by region (hyperbolic FDH; $p = q = 1$)

Region	Statistic	p-value
1	0.820	3.875(−74)
2	0.916	1.687(−8)
3	0.903	1.914(−91)
4	0.967	9.199(−1)

Note: Values in parentheses give value of exponent in scientific notation; for example, 3.875×10^{-74} is denoted by 3.875(−74).

Estimation of the technical change measure defined in (4.40) requires estimation of distances from observations in one period to the frontier in another period as revealed by the expression in (4.40). To avoid numerical problems arising from not being able to reach the frontier from a different period in the input or output directions, we estimate the technical change measure using only the hyperbolic direction. The results in Table 4.9 reveal statistics that are significantly less than one for each of the four regions, indicating that technology shifted downward from 2002 to 2012. The separability tests in Table 4.5 indicate that the technology is *different* in 2002 and 2012 for each region, but the results here give evidence on the *direction* of the change. For region 1, the improvement in efficiency suggested by results in Table 4.10 is offset to some degree by the downward shift of the efficient frontier. While municipalities in region 1 may have increased their efficiency, this was due in part to the downward shift in the technology as opposed to an increase in productivity.

8. CONCLUSIONS

The discussions of techniques and methods in this chapter are, for purposes of brevity, limited to radial (that is, proportional) efficiency measures. However, all of the results discussed above extend to directional efficiency measures developed by Chambers et al. (1996, 1998). With directional measures, efficiency of a unit is measured by its distance to the efficient boundary in a direction fixed by a given vector in $\mathbb{R}_{+}^{p+q}$; efficiency is given by an additive quantity as opposed to a multiplicative factor as in the case of radial measures, but otherwise the details and associated statistical issues are similar. Asymptotic properties of FDH estimators of these directional distances are developed by Simar and Vanhems (2012), and by Simar et al. (2012) for corresponding DEA estimators. The CLTs developed by Kneip et al. (2015b) and Daraio et al. (2018) as well as the tests developed by Kneip et al. (2016) and Daraio et al. (2018) described above extend easily to additive, directional distances.

It is well known that envelopment estimators such as those discussed in this chapter are susceptible to outliers in the data. As discussed by Wilson (1993) and others, outliers can be observations from the tail of a distribution, and hence important to consider, or they can result from corrupted data in which case they should be removed. Various methods for detecting outliers have been developed, including those described by Wilson (1993, 1995), Simar (2003) and Porembski et al. (2005). Finding outliers in more than a few dimensions is a formidable task, and more than one method should be used. Alternatively, one can use robust (with respect to outliers) methods based on order-α frontiers (for example, see Aragon et al., 2005; Daouia and Simar, 2007; Wheelock and Wilson, 2008) or order-*m* frontiers (for example, see Cazals et al., 2002). Both of these ideas are based on the idea of estimating not the full frontier Ψ^{∂}, but rather a feature (that is, a 'partial frontier') lying close to the full frontier. Order-α frontiers are based on quantiles, while order-*m* frontiers use a trimming parameter (denoted by *m*) to eliminate the influence of potential outliers or extreme points in the data. Both methods provide benchmarks against which efficiency can be measured. See Simar and Wilson (2013) for additional discussion.

Some research also investigates how noise in the observation process might be accounted for, along the lines of the so-called stochastic frontier analysis, but without introducing restrictive functional-form assumptions. The literature on stochastic frontier analysis involves fully parametric approaches, but this new strain in the literature focuses on nonparametric and semiparametric models and methods. So far, most of the work in this area is limited to univariate response models (for example, production functions or cost functions); for example, see the work by Kumbhakar et al. (2007), Kneip et al. (2015a), Simar et al. (2017) and Florens et al. (2020). Multivariate extensions include Simar (2007), Simar and Zelenyuk (2011) and Simar and Wilson (2023).

For too many years, FDH and DEA estimators were used without any statistical underpinnings. Papers in the literature from the 1980s and beyond often imply that truth can be learned from data, but in fact, one can only obtain estimates and make inferences using data. Today, however, empirical researchers can bring a 'tool box' full of statistical tools to data in order to conduct rigorous analyses and to quantify how the uncertainty surrounding estimates can be quantified. The various testing methods discussed here have been implemented in the FEAR software package introduced by Wilson (2008), and so there is little excuse for not using these methods.

NOTES

1. The review by Narbón-Perpiñá and De Witte (2018, Table A2) lists 97 empirical studies of local government efficiency. Sixty-six of these studies use nonparametric estimators, while only 31 use parametric methods. Among the 66 papers employing nonparametric estimators, 50 use DEA estimators, 14 use FDH estimators, and 2 use both.
2. Free disposability means that for any $\tilde{x} \geq x$ and $\tilde{y} \leq y$, if (x, y) $\in \Psi$ then $(\tilde{x}, \tilde{y}) \in \Psi$.
3. As seen below, methods exist for reducing the dimensionality of a given problem. In many cases, this substantially reduces estimation error.
4. See also the recent surveys by Bădin et al. (2012, 2014).
5. Unfortunately, papers continue to appear in which these points are ignored. Such papers should be disregarded.
6. The separability test with discrete environmental variables appears in Daraio et al. (2018, Appendix C).
7. Alternatively, one can stratify the splitting, by splitting each subgroups of data χ_n^k independently for $k = 1, \ldots, K$.
8. Separate versions of the CPI are available for select cities. Where these are not available, we use regional CPIs to adjust for price differences.
9. A map showing census regions of the United States is available at https://www2.census.gov/programs-surveys/sahie/reference-maps/2014/us-regdiv.jpg.

REFERENCES

Adler, N. and B. Golany (2007), PCA-DEA: Reducing the curse of dimensionality, in J. Zhu and W. D. Cook, eds, *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis*, New York: Springer Science+Business Media, LLC, pp. 139–154.

Afonso, F. S. (2008), Assessing and explaining the relative efficiency of local government, *Journal of Socio-Economics* 37, 1946–1979.

Afriat, S. (1972), Efficiency estimation of production functions, *International Economic Review* 13, 568–598.

Aigner, D., C. A. K. Lovell, and P. Schmidt (1977), Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* 6, 21–37.

Aragon, Y., A. Daouia, and C. Thomas-Agnan (2005), Nonparametric frontier estimation: A conditional quantile-based approach, *Econometric Theory* 21, 358–389.

Bădin, L., C. Daraio, and L. Simar (2012), How to measure the impact of environmental factors in a nonparametric production model, *European Journal of Operational Research* 223, 818–833.

——— (2014), Explaining inefficiency in nonparametric production models: The state of the art, *Annals of Operations Research* 214, 5–30.

——— (2019), A bootstrap approach for bandwidth selection in estimating conditional efficiency measures, *European Journal of Operational Research* 277, 784–797.

Banker, R. D., A. Charnes, and W. W. Cooper (1984), Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science* 30, 1078–1092.

Cazals, C., J. P. Florens, and L. Simar (2002), Nonparametric frontier estimation: A robust approach, *Journal of Econometrics* 106, 1–25.

Chambers, R. G., Y. Chung, and R. Färe (1996), Benefit and distance functions, *Journal of Economic Theory* 70, 407–419.

——— (1998), Profit, directional distance functions, and nerlovian efficiency, *Journal of Optimization Theory and Applications* 98, 351–364.

Charnes, A., W. W. Cooper, and E. Rhodes (1978), Measuring the efficiency of decision making units, *European Journal of Operational Research* 2, 429–444.

Da Cruz, N. F. and R. C. Marques (2014), Revisiting the determinants of local government performance, *Omega* 44, 91–103.

Daouia, A. and L. Simar (2007), Nonparametric efficiency analysis: A multivariate conditional quantile approach, *Journal of Econometrics* 140, 375–400.

Daouia, A., L. Simar, and P. W. Wilson (2017), Measuring firm performance using nonparametric quantile-type distances, *Econometric Reviews* 36, 156–181.

Daraio, C. and L. Simar (2005), Introducing environmental variables in nonparametric frontier models: A probabilistic approach, *Journal of Productivity Analysis* 24, 93–121.

——— (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis*, New York: Springer Science+Business Media, LLC.

Daraio, C., L. Simar, and P. W. Wilson (2018), Central limit theorems for conditional efficiency measures and tests of the 'separability condition' in non-parametric, two-stage models of production, *The Econometrics Journal* 21, 170–191.

De Borger, B. and K. Kerstens (2000), What is known about municipal efficiency? the Belgian case and beyond, in J. Blank, ed., *Public Provision and Performance: Contributions from Efficiency and Productivity Measurement*, Amsterdam: Elsevier, pp. 299–330.

de Oliveira Junqueira, M. (2015), Efficiency measurement of local public sector in international perspective: A literature survey. Unpublished working paper, University of São Paulo, Brazil. Presented at International Conference on Public Policy, Milan, 2015.

Debreu, G. (1951), The coefficient of resource utilization, *Econometrica* 19, 273–292.

Deprins, D., L. Simar, and H. Tulkens (1984), Measuring labor inefficiency in post offices, in M. M. P. Pestieau and H. Tulkens, eds, *The Performance of Public Enterprises: Concepts and Measurements*, Amsterdam: North-Holland, pp. 243–267.

Ellickson, B. (1971), Jurisdicitonal fragmentation and residential choice, *American Economic Review* 61, 334–339.

Färe, R., S. Grosskopf, B. Lindgren, and P. Roos (1992), Productivity changes in Swedish pharmacies 1980–1989: A non-parametric Malmquist approach, *Journal of Productivity Analysis* 3, 85–101.

Färe, R., S. Grosskopf, and C. A. K. Lovell (1985), *The Measurement of Efficiency of Production*, Boston: Kluwer-Nijhoff Publishing.

Farrell, M. J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society A* 120, 253–281.

Florens, J. P., L. Simar, and I. Van Keilegom (2020), Estimation of the boundary of a variable observed with symmetric error, *Journal of the American Statistical Association* 115, 425–441.
Grosskopf, S., K. J. Hayes, L. L. Taylor, and W. L. Weber (2001), On the determinants of school district efficiency: Competition and monitoring, *Journal of Urban Economics* 49, 453–478.
Grossman, P. J., P. Mavros, and R. W. Wassmer (1999), Public sector technical inefficiency in large U.S. cities, *Journal of Urban Economics* 46, 278–299.
Härdle, W. and L. Simar (2019), *Applied Multivariate Statisticial Analyisis*, Berlin: Springer-Verlag. Fifth edition.
Hayes, K. J., L. Razzolini, and L. B. Ross (1998), Bureaucratic choice and nonoptimal provison of public goods: Theory and evidence, *Public Choice* 94, 1–20.
Henderson, J. M. and R. E. Quandt (1971), *Microeconomic Theory: A Mathematical Approach*, New York: McGraw-Hill Book Company.
Jeong, S. O., B. U. Park, and L. Simar (2010), Nonparametric conditional efficiency measures: Asymptotic properties, *Annals of Operations Research* 173, 105–122.
Jeong, S. O. and L. Simar (2006), Linearly interpolated FDH efficiency score for nonconvex frontiers, *Journal of Multivariate Analysis* 97, 2141–2161.
Kneip, A., B. Park, and L. Simar (1998), A note on the convergence of nonparametric DEA efficiency measures, *Econometric Theory* 14, 783–793.
Kneip, A., L. Simar, and I. Van Keilegom (2015a), Frontier estimation in the presence of measurement error with unknown variance, *Journal of Econometrics* 184, 379–393.
Kneip, A., L. Simar, and P. W. Wilson (2008), Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models, *Econometric Theory* 24, 1663–1697.
——— (2011), A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators, *Computational Economics* 38, 483–515.
——— (2015b), When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores, *Econometric Theory* 31, 394–422.
——— (2016), Testing hypotheses in nonparametric models of production, *Journal of Business and Economic Statistics* 34, 435–456.
——— (2021), Inference in dynamic, nonparametric models of production: Central limit theorems for Malmquist indices, *Econometric Theory* 37, 537–572.
Koopmans, T. C. (1951), An analysis of production as an efficient combination of activities, in T. C. Koopmans, ed., *Activity Analysis of Production and Allocation*, New York: John-Wiley and Sons, Inc., pp. 33–97. Cowles Commission for Research in Economics, Monograph 13.
Kumbhakar, S. C., B. U. Park, L. Simar, and E. G. Tsionas (2007), Nonparametric stochastic frontiers: A local likelihood approach, *Journal of Econometrics* 137, 1–27.
Lesourne, J. (1972), *Le Calcul Économique*, Paris: Dunod.
lo Storto, C. (2013), Evaluating technical efficiency of Italian major municipalities: A data envelopment analysis model, *Procedia-Social and Behavioral Sciences* 81, 346–350.
Mas-Colell, A., M. D. Whiston, and J. R. Green (1995), *Microeconomic Theory*, New York: Oxford University Press, Inc.
Meeusen, W. and J. van den Broeck (1977), Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* 18, 435–444.
Narbón-Perpiñá, I. and K. De Witte (2018), Local governments' efficiency: a systematic literature review—part I, *International Transactions in Operational Research* 25, 431–468.
Park, B. U., S.-O. Jeong, and L. Simar (2010), Asymptotic distribution of conical-hull estimators of directional edges, *Annals of Statistics* 38, 1320–1340.
Park, B. U., L. Simar, and C. Weiner (2000), FDH efficiency scores from a stochastic point of view, *Econometric Theory* 16, 855–877.
Porembski, M., K. Breitenstein, and P. Alpar (2005), Visualizing efficiency and reference relations in data envelopment analysis with an application to the branches of a German bank, *Journal of Productivity Analysis* 23, 203–221.
Robinson, P. M. (1988), Root-*n*-consistent semiparametric regression, *Econometrica* 56, 931–954.
Simar, L. (2003), Detecting outliers in frontier models: A simple approach, *Journal of Productivity Analysis* 20, 391–424.
——— (2007), How to improve the performances of DEA/FDH estimators in the presence of noise, *Journal of Productivity Analysis* 28, 183–201.
Simar, L., I. Van Keilegom, and V. Zelenyuk (2017), Nonparametric least squares methods for stochastic frontier models, *Journal of Productivity Analysis* 47, 189–204.
Simar, L. and A. Vanhems (2012), Probabilistic characterization of directional distances and their robust versions, *Journal of Econometrics* 166, 342–354.
Simar, L., A. Vanhems, and P. W. Wilson (2012), Statistical inference for DEA estimators of directional distances, *European Journal of Operational Research* 220, 853–864.

Simar, L. and P. W. Wilson (2007), Estimation and inference in two-stage, semi-parametric models of productive efficiency, *Journal of Econometrics* 136, 31–64.
——— (2011a), Inference by the *m* out of *n* bootstrap in nonparametric frontier models, *Journal of Productivity Analysis* 36, 33–53.
——— (2011b), Two-Stage DEA: Caveat emptor, *Journal of Productivity Analysis* 36, 205–218.
——— (2013), Estimation and inference in nonparametric frontier models: Recent developments and perspectives, *Foundations and Trends in Econometrics* 5, 183–337.
——— (2015), Statistical approaches for non-parametric frontier models: A guided tour, *International Statistical Review* 83, 77–110.
——— (2019), Central limit theorems and inference for sources of productivity change measured by nonparametric Malmquist indices, *European Journal of Operational Research* 277, 756–769.
——— (2020a), Hypothesis testing in nonparametric models of production using multiple sample splits, *Journal of Productivity Analysis* 53, 287–303.
——— (2020b), Technical, allocative and overall efficiency: Estimation and inference, *European Journal of Operational Research* 282, 1164–1176.
——— (2023), Nonparametric, stochastic frontier models with multiple inputs and outputs. Journal of Business and Economic Statistics, forthcoming.
Simar, L. and V. Zelenyuk (2011), Stochastic FDH/DEA estimators for frontier analysis, *Journal of Productivity Analysis* 36, 1–20.
Štastná, L. and M. Gregor (2015), Public sector efficiency in transition and beyond: Evidence from Czech local governments, *Applied Economics* 47, 680–699.
Tang, K.-L. (1997), Efficiency in the private sector: A critical review of empirical evidence from public services, *International Review of Administrative Sciences* 63, 459–474.
Tiebout, C. M. (1956), A pure theory of local public expenditure, *Journal of Political Economy* 64, 416–424.
Tran, C.-D. T. T. and B. Dollery (2020), An empirical analysis of municipal efficiency and local resident satisfaction by council type in Victorian local government, *Australian Journal of Public Administration* 79, 531–549.
Varian, H. R. (1978), *Microeconomic Analysis*, New York: W. W. Norton & Co.
Wheelock, D. C. and P. W. Wilson (2008), Non-parametric, unconditional quantile estimation for efficiency analysis with an application to Federal Reserve check processing operations, *Journal of Econometrics* 145, 209–225.
Wilson, P. W. (1993), Detecting outliers in deterministic nonparametric frontier models with multiple outputs, *Journal of Business and Economic Statistics* 11, 319–323.
——— (1995), Detecting influential observations in data envelopment analysis, *Journal of Productivity Analysis* 6, 27–45.
——— (2008), FEAR: A software package for frontier efficiency analysis with R, *Socio-Economic Planning Sciences* 42, 247–254.
——— (2011), Asymptotic properties of some non-parametric hyperbolic efficiency estimators, in I. Van Keilegom and P. W. Wilson, eds., *Exploring Research Frontiers in Contemporary Statistics and Econometrics*, Berlin: Springer-Verlag, pp. 115–150.
——— (2018), Dimension reduction in nonparametric models of production, *European Journal of Operational Research* 267, 349–367.

PART II

PUBLIC SECTOR EFFICIENCY, AND INSTITUTIONS AND FIRMS

5. Overall public sector efficiency

Antonis Adam and Sofia Tsarsitalidou

1. INTRODUCTION

Evaluating overall Public Sector Efficiency (PSE) is crucial for policy analysis. Despite the issue's importance, there are still not many studies examining overall PSE at the country level. Even though Afonso et al. (2005) have provided a framework for aggregating different government activities in a single index, data limitations do not allow the use of the index in a large cross-section of countries. Moreover, most non-parametric methods (Data Envelopment Analysis or Free Disposable Hull) used to evaluate how public sector inputs are translated into public sector performance require a homogeneous sample of countries.[1] Hence, most of the existing studies in overall PSE typically cover only high-income countries.

In the present chapter, we construct PSE scores for many countries, developing and developed, and then examine the external factors that affect the efficiency of the public sector. Our initial step is to briefly present previous studies that measure overall PSE and the main external variables found in this literature as determinants of PSE. The main focus is on country-level studies.[2] Then, as a next step, we compile PSE indicators, using partial frontier efficiency analysis and, more precisely, order-*m* efficiency scores (Tauchmann 2012). The advantage of this methodology is that it generalizes Free Disposable Hull (FDH) by adding a layer of randomness to PSE computation. By computing the efficiency scores *D* times, for *m* units each time, (i) it allows observations to lie outside the efficiency frontier, (ii) it makes the frontier less vulnerable to outliers and (iii) the results do not suffer from the curse of dimensionality (Daraio and Simar 2007).

To measure the effect of external factors, we follow the method of Daraio and Simar (2007) and De Witte and Kortelainen (2013). This method uses the share of conditional to unconditional order-*m* scores and employs non-parametric regression methods to determine the influence of environmental variables. This method overcomes several disadvantages of the Ordinary Least Squares, tobit or double bootstrap analysis (Simar and Wilson 2007), which were typically used to assess the effect of external factors.[3] Furthermore, our results verify several previous findings regarding the external environment's effect, at least qualitatively.

2. MEASURING PSE

2.1 Previous Studies

Several contributions have tried to establish a standard measure of overall PSE. The seminal study of Afonso et al. (2005) constructs measures of PSE for 23 industrialized countries over the 1990–2000 period, using a FDH analysis. To measure public sector performance (PSP), this paper defines seven areas of government activity and combines them as a performance indicator. These are sub-indicators of administrative, education, health and public infrastructure outcomes, economic stability, income distribution and allocative efficiency. The former four sub-indicators are what the authors define as 'opportunity indicators', whereas the latter three cover the Musgravian tasks of the government.

Specifically, for country *i* and activity area *j*

$$PSP_i = \sum_{j=1}^{n} PSP_{ij}$$

where $PSP_{ij} = f(I_k)$, with *k* the number of economic and social indicators of each area of government activity. The PSP_{ij} is computed and normalized over the country sample average. The overall PSP score is simply the sum of all performance indicators. The overall PSP indicator is then used as the FDH analysis output, where the related input is considered the overall level of public sector spending. As a next step, the authors employ FDH analysis to rank the efficiency of producers. Thus, they compare each individual performance with a theoretical production possibility frontier. With this analysis, they can determine the lowest level of inputs necessary to produce a given level of outputs, when they compute input-oriented efficiency scores, and the highest level of output for a given level of inputs, when the output-oriented scores are calculated. Hence, a measure of the slack of public expenditures is constructed, and the relative efficiency of each country's public sector is evaluated.

The construction of an overall PSP indicator, to our understanding, has two fundamental advantages. First, it allows the aggregation of all government outputs into a single index, and second, it deals with the dimensionality curse (Park et al. 2000) when applying the FDH approach.

Several papers have followed the tradition of Afonso et al. (2005) to derive measures of PSE. This literature is highlighted in Table 5.1. The common element in these papers is that first, performance indicators for each area of public sector activity are derived. Then, these separate scores are aggregated to a single PSP measure which is later used as an output indicator in a parametric or non-parametric analysis.

Despite these standard features, there is high heterogeneity across studies on the country sample employed, the socio-economic indicators considered to be the output of the public sector and the method employed to estimate the public sector slack. As the table suggests, most studies on overall PSE use a sample of developed countries. The exceptions to this are Herrera and Pang (2005), Angelopoulos et al. (2008), Hauner and Kyobe (2010) and Adam et al. (2011a). Even though these studies have a larger number of countries with different public production technologies and different institutional environments, their PSP analysis is limited by the availability of data. They, thus, resort to a more restricted array of public sector activities, as well as use fewer output indicators in each area.

Regarding the assumed socio-economic outputs of the public sector, there is a trade-off between country/year coverage of the PSE scores and the quality of the index. Some studies limit their analysis to Education and Health (for example, Herrera and Pang 2005; Hauner and Kyobe 2010; Adam et al. 2014), where both the effect of private sector involvement and the quality of the final output can be taken into account. Other studies (most notably Christl et al. 2020) use a variety of public sector output indicators.

Finally, most studies use non-parametric techniques, that is, FDH and DEA techniques (Herrera and Pang 2005; Hauner and Kyobe 2010; Afonso et al. 2005, 2010; Adam et al. 2011b,

Table 5.1 Summary of studies

Study	Outputs	Inputs	Countries	Method
Afonso et al. (2005)	Administrative: corruption, red tape, quality of the judiciary, size of the shadow economy. Education: secondary enrollment, OECD educational attainment index. Health: infant mortality and life expectancy. Public Infrastructure: quality of communication and transport infrastructure. Economic Stability: stability of the output growth, average inflation. Income Distribution: share of the poorest 40% households. Allocative Efficiency: GDP per capita, GDP growth, unemployment.	Total public spending as a percent of GDP. Spending on goods and services (administration). Spending on health and education. Social payments. Public investment. Total public spending (economic performance).	Industrialized Countries	Public Sector Performance FDH
Afonso et al. (2010)	As in Afonso et al. (2005).		Industrialized Countries	DEA
Afonso and Aubyn (2013)	Human Capital. Public Physical Capital. Private Physical Capital.	GDP per worker.	OECD countries	DEA Malmquist productivity Stochastic Frontier Analysis (SFA)
Angelopoulos et al. (2008)	Education: secondary school enrollment. Administration: (absence of) corruption and bureaucratic quality. Infrastructure: Diesel locomotives in use, (inverse of) power transmission and distribution losses. Stabilization: Inverse of inflation and unemployment.	Public Expenditure education in each policy area. Total Government Expenditure (infrastructure).	64 countries (high and lower income)	Public Sector Performance SFA

(Continued)

Table 5.1 (Continued)

Study	Outputs	Inputs	Countries	Method
Papaioannou (2016)	Education: tertiary school enrollment. Administration: (absence of) corruption, judicial independence, protection of property rights. Infrastructure: volume of freight carried by air transport, fixed telephone subscriptions, mobile telephone users, number of internet users. Stabilization: Inverse of inflation and unemployment.	Public consumption as share in GDP (administration and stabilization). Public spending on education as a percentage of GDP. Public investment as a share of GDP (infrastructure).	EU 28	
Adam et al. (2014)	Education: years of schooling multiplied by education quality indicator and by share of public to total educational spending. Health: life expectancy at birth and inverse of infant mortality rate both multiplied by the share of public to total spending on health.	Public spending on education as a percentage of GDP. Public spending on health as a percentage of GDP.	OECD countries	SFA
Adam et al. (2011b)	Education: secondary school enrollment multiplied by education quality indicator and by share of public to total educational spending. Health: life expectancy at birth and inverse of infant mortality rate both multiplied by the share of public to total spending on health. Social Security and Welfare: inverse of income inequality. General Public Services: (Inverse of) corruption, bureaucratic quality. Economic Affairs: (inverse of) power transmission and distribution losses, fixed telephone lines access.	Public spending on education. Public spending on health. Public spending on social security and welfare. Total public spending (Economic Affairs and General Public Services).	OECD countries	DEA SFA
Adam et al. (2011a)	As in Angelopoulos et al. (2008)		64 countries (high and lower income)	Public Sector Performance

Table 5.1 (Continued)

Study	Outputs	Inputs	Countries	Method
Christl et al. (2020)	Administration: quality of public services, independence of the judiciary, absolute legal institutional quality, level of the shadow economy, corruption perceptions index, public-sector corruption index, regulatory quality. Education: use labor force with primary education, quality of primary education, quality of the educational system, young people not in education or employment, pupil-teacher ratio in primary, secondary and tertiary education, Program for International Student Assessment score. Health: life expectancy, healthy life years at birth for males and females, (inverse of) infant mortality. Public Security: inverse of crime, violence, or vandalism in the area, organized crime, and reliability of police services. Environmental Protection: air quality, environmental health, environmental performance index. Social Security: pension generosity index, poverty gap, poverty rate 50%. Defense: armed forces personnel. Infrastructure: quality of overall infrastructure. Economic Indicators: general government net lending, GDP per capita growth, general government debt, real GDP growth. Stability: stability of economic growth, stability of inflation, political stability, and general government debt. Inequality: Gini coefficient.	Public sector expenditure for each category. Administration: General public services expenditure. Health: Health expenditure. Environmental Protection: Environment protection expenditure. Public security: Public order and safety expenditure. Social security: Social protection expenditure. Defense: Defense expenditure. Infrastructure: Total inland transport infrastructure investment. Total: Total expenditure.	23 European Countries	FDH order-m

(Continued)

Table 5.1 (Continued)

Study	Outputs	Inputs	Countries	Method
Herrera and Pang (2005)	Education: primary school enrollment (gross and net), secondary school enrollment (gross and net), literacy of youth, average years of school, first level complete, second-level complete, and learning scores. Health: life expectancy at birth, DPT and measles immunization, disability-adjusted life expectancy.	Public spending on Education. Public spending on Health.	140 Developed and Developing	DEA FDH
Hauner and Kyobe (2010)	Education: Primary and secondary enrolment rates, primary education completion rate. Health: DPT immunization rate, infant mortality rate, mortality rate of female adults.	General Government Spending Education. General Government Spending Health.	114 Developed and Developing	

Source: Authors.

2014; Afonso and Aubyn 2013; Christl et al. 2020), whereas some others use only parametric, for example, stochastic frontier analysis (Afonso and Aubyn 2013; Angelopoulos et al. 2008).

Apart from the studies measuring the PSE in a cross-country framework, many studies focused on local PSE in provinces or municipalities inside a specific country. According to Afonso and Fernandes (2008), examining the efficiency of the local governments is essential for various reasons, as the comparison across similar units is possible, and relative efficiency can be evaluated more efficiently. Since the present paper deals with overall public sector efficiency in a cross-country context, the interested reader is referred to the recent and extensive survey of Narbón-Perpiñá and De Witte (2018a, 2018b) for more details on this.

2.2 Measurement

The existing method of constructing a single PSP measure implicitly assumes that the different areas of public sector activity are perfect substitutes. For example, a country may exhibit extremely low efficiency in the education area and be highly efficient in its health policy. Then, the aggregation of the two sub-indicators might make one country appear more efficient than one with an (above) average efficiency in both policy areas. This is particularly important because the typical input measures used to calculate efficiency scores are all part of the overall government budget constraint. Hence, an inefficiently high allocation in one policy activity is ‘compensated’ by an inefficiently low allocation in another, giving rise to a very high-efficiency score. To make matters worse, DEA and FDH approaches are sensitive to outliers. Hence, the pooling of high-income and low-income countries might be problematic.[4]

To deal with these issues, in the present section, we employ a partial frontier approach (for example, Daraio and Simar 2007; Tauchmann 2012) and, more specifically, an order-*m* efficiency approach (Cazals et al. 2002) to compute input-oriented efficiency scores. Order-*m* efficiency extends the FDH model, which adds a layer of randomness. Specifically, a sample of *m* decision-making units (DMUs) is drawn with replacement from the overall sample. Then, a pseudo-FDH efficiency score is calculated for this sub-sample, and the whole process is repeated a large number of times, *D*.[5] The final efficiency score is the average of the pseudo-FDH scores. This method has several advantages. First, it does not impose a maximum value of 1 on the efficiency score, that is, it allows for super-efficient DMUs. Second, the random resampling deals with the problem of outliers and measurement error, as the final order-*m* efficiency score is not sensitive to the existence of particular outliers in the whole sample. Finally, given that the entire measure is computed *D* times, there is no 'curse of dimensionality' (Daraio and Simar 2007).

Given the above, the order-*m* efficiency scores might be ideal for a large cross-section of countries. First, with the inclusion of developing countries in the sample, we should expect the presence of both outliers and measurement error. Moreover, as we want to measure overall PSE, we do not want to impose the substitutability of outputs by computing an overall PSP index.

The following table presents the input and outputs used in the analysis. All data are taken from the Quality of Governance (QoG) project (Teorell et al. 2020). We follow the literature highlighted in the previous sub-section and define public sector performance in five areas: public administration, education, health, infrastructure and government stability. These are the policy areas included in most studies on overall PSE. As our focus is on providing a measure for as many countries as possible, we used one variable for each area of government activity. For that reason, we have tried to find the most general variable for each category. The indicators, definitions and data sources for output and input indicators are presented in Table 5.2.

Table 5.2 Variables employed in the PSE indicators

Indicator	Definition	Source
Public Sector Performance indicators (outputs)		
Administration	ICRG synthetic index of bureaucracy, corruption and law and order	QoG- ICRG country risk guide
Education	Gross secondary school enrollment	QoG- World Bank's World Development Indicators (WDI)
Health	The inverse of infant mortality (per 1000 births)	QoG- WDI
Infrastructure	Number of telephone lines (per 100 inhabitants)	QoG- WDI
Stability	100-Unemployment rate	QoG- WDI
Public Sector Expenditures (input)		
Total Public Sector Expenditures	Government consumption as a share of GDP	QoG- WDI

Source: Authors

Furthermore, we have used two time periods, namely 2000–10 and 2011–20, taking averages of the corresponding variables. This ensures that short-run fluctuations are canceled out. Moreover, we expect that the technology of the public sector does not change within the same time period.

Table 5.3 presents the countries' input-oriented efficiency scores, that is, the order-*m* and the FDH indexes and the simple DEA score. Moreover, we also present the relative rankings of countries for the two associated periods. Following Daraio and Simar (2005), we choose *m* so that the number of super-efficient observations is relatively constant.[6] The results are also depicted in Figure 5.1 and Figure 5.2.

Table 5.3 Efficiency Scores

	2001–10				2011–20			
Country	Order-*m*	Ranking	FDH	DEA	Order-*m*	Ranking	FDH	DEA
Albania	0.973	54	0.849	0.524	0.774	62	0.761	0.547
Algeria	0.758	85	0.662	0.402	0.499	97	0.433	0.317
Angola	0.345	119	0.284	0.213	0.535	89	0.347	0.267
Argentina	0.892	66	0.803	0.568	0.586	80	0.586	0.457
Austria	1.000	19	0.602	0.525	1.000	5	0.605	0.519
Bahamas	1.001	12	1.000	0.856	0.779	60	0.777	0.681
Bahrain	1.000	19	0.728	0.523	1.000	5	0.611	0.479
Bangladesh	1.001	6	1.000	0.913	1.000	1	1.000	0.904
Armenia	1.010	3	1.000	0.665	0.812	56	0.701	0.534
Belgium	1.000	19	0.519	0.491	1.000	5	0.598	0.463
Bolivia	0.961	56	0.666	0.397	0.731	66	0.536	0.380
Botswana	0.522	110	0.493	0.296				
Brazil	0.891	67	0.552	0.375	0.503	96	0.503	0.367
Brunei	0.519	111	0.470	0.352	0.430	103	0.430	0.340
Bulgaria	0.644	96	0.581	0.433	0.601	77	0.601	0.459
Myanmar	1.000	19	0.476	0.415	0.924	50	0.315	0.269
Belarus	0.888	68	0.550	0.438	1.000	5	0.670	0.643
Cameroon	0.837	72	0.462	0.382	0.926	49	0.473	0.414
Canada	1.000	19	0.583	0.562	1.000	5	0.570	0.511
Sri Lanka	1.000	14	0.750	0.488	1.000	5	1.000	0.801
Chile	1.000	13	0.930	0.691	0.744	65	0.744	0.597

Table 5.3 *(Continued)*

	2001–10				2011–20			
Country	Order-*m*	Ranking	FDH	DEA	Order-*m*	Ranking	FDH	DEA
China	0.774	82	0.700	0.469				
Colombia	0.727	89	0.687	0.449	0.681	70	0.590	0.461
Congo	0.441	116	0.359	0.290	0.496	99	0.441	0.341
Congo, DR	1.000	15	0.813	0.590	0.730	67	0.726	0.496
Costa Rica	0.811	76	0.734	0.538	1.000	5	0.569	0.435
Croatia	0.610	100	0.609	0.478	0.593	79	0.490	0.449
Cuba	1.000	19	0.287	0.184	1.000	5	0.266	0.207
Cyprus	1.000	19	0.680	0.622	1.000	5	0.585	0.574
Czech Republic	0.798	77	0.512	0.405	0.498	98	0.498	0.393
Denmark	1.000	19	0.621	0.465	1.000	5	0.469	0.396
Dominican Republic	1.001	8	1.000	0.619	0.826	54	0.821	0.569
Ecuador	0.961	57	0.897	0.547	1.000	5	0.695	0.486
El Salvador	0.744	88	0.694	0.427	0.612	74	0.532	0.392
Estonia	0.938	59	0.581	0.475	1.000	5	0.505	0.451
Finland	1.000	19	0.540	0.473	1.000	5	0.415	0.362
France	1.000	19	0.506	0.464	1.000	5	0.597	0.482
Gabon	0.602	103	0.355	0.318				
Gambia	1.012	2	0.900	0.564				
Germany	1.000	19	0.616	0.607	1.000	5	0.720	0.583
Ghana	0.964	55	0.518	0.469	0.896	52	0.825	0.505
Greece	0.931	61	0.574	0.486	1.000	5	0.585	0.489
Guatemala	1.000	19	0.887	0.551	1.000	5	0.791	0.536
Guinea	1.001	11	0.650	0.548	0.470	102	0.268	0.211
Guyana	0.505	114	0.503	0.321	0.597	78	0.596	0.428
Honduras	0.670	93	0.515	0.336	0.577	83	0.577	0.353
Hungary	0.780	81	0.483	0.396	0.491	100	0.491	0.433
Iceland	1.000	19	0.654	0.493	1.000	5	0.592	0.473
India	1.000	19	0.770	0.517	1.000	5	0.794	0.540
Indonesia	1.013	1	1.000	0.679	1.000	3	0.922	0.677

(Continued)

Table 5.3 (Continued)

	2001–10				2011–20			
Country	Order-*m*	Ranking	FDH	DEA	Order-*m*	Ranking	FDH	DEA
Iran	0.933	60	0.930	0.654	1.000	2	0.881	0.759
Iraq	0.458	115	0.280	0.238				
Ireland	1.000	19	0.684	0.616	1.000	5	0.828	0.727
Israel	0.701	91	0.480	0.409	1.000	5	0.525	0.447
Italy	0.973	53	0.600	0.469	0.509	95	0.509	0.453
Jamaica	0.716	90	0.678	0.436	0.691	68	0.591	0.428
Kazakhstan	1.001	9	0.932	0.628	1.000	5	0.939	0.735
Jordan	0.515	113	0.487	0.308	0.631	73	0.539	0.354
Kenya	0.532	108	0.321	0.285				
Korea, South	1.000	19	0.892	0.757	1.000	5	0.774	0.709
Kuwait	1.000	19	0.572	0.417	0.579	81	0.488	0.352
Latvia	0.885	69	0.548	0.422	0.803	58	0.549	0.437
Libya	1.003	5	0.770	0.503				
Lithuania	0.871	70	0.540	0.402	0.577	82	0.577	0.450
Luxembourg	1.000	19	0.709	0.660	1.000	5	0.845	0.654
Madagascar	0.762	84	0.322	0.275	1.000	5	0.373	0.294
Malawi	0.792	78	0.423	0.369	0.604	76	0.396	0.329
Malaysia	1.000	19	0.865	0.577	0.907	51	0.765	0.560
Mali	0.394	118	0.348	0.267	0.531	91	0.342	0.255
Malta	1.000	19	0.607	0.534	1.000	5	0.667	0.599
Mexico	1.000	18	0.981	0.636	0.824	55	0.824	0.577
Mongolia	0.833	73	0.717	0.435				
Moldova	0.620	98	0.561	0.382	0.649	72	0.649	0.511
Morocco	0.594	104	0.520	0.320	0.514	94	0.442	0.313
Mozambique	0.530	109	0.303	0.230	0.414	105	0.239	0.183
Oman	0.577	105	0.497	0.331	1.000	5	0.398	0.280
Netherlands	1.000	19	0.500	0.462	1.000	5	0.474	0.428
New Zealand	1.000	19	0.642	0.578	1.000	5	0.636	0.558
Nicaragua	0.919	62	0.707	0.465				
Niger	0.609	101	0.290	0.205	1.000	5	0.345	0.248

Table 5.3 (Continued)

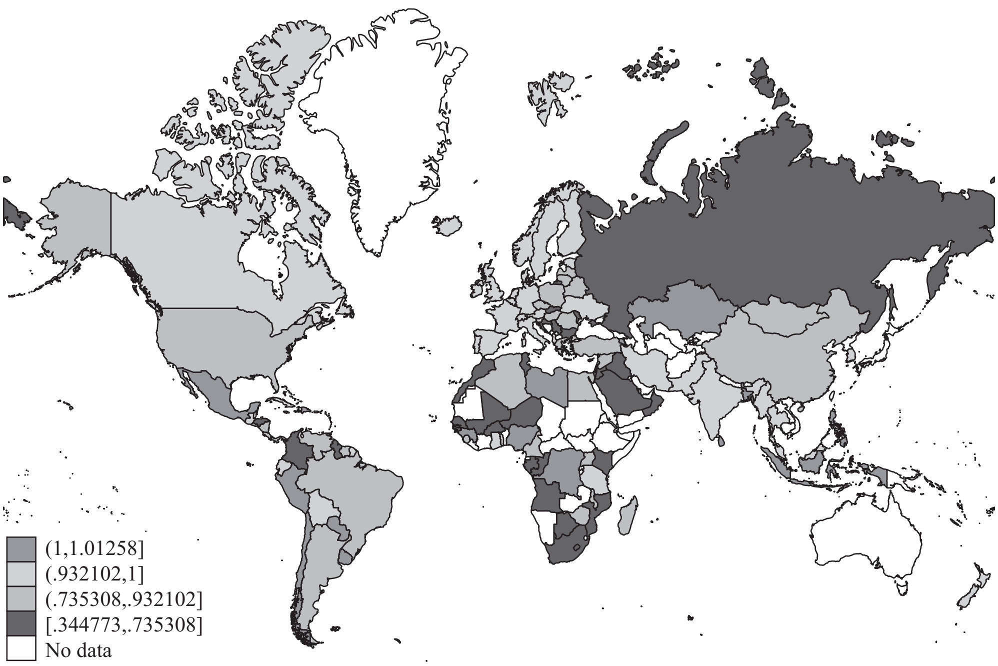
	2001–10				2011–20			
Country	Order-*m*	Ranking	FDH	DEA	Order-*m*	Ranking	FDH	DEA
Nigeria	1.001	10	0.934	0.737	0.873	53	0.868	0.659
Norway	1.000	19	0.574	0.515	1.000	5	0.433	0.393
Pakistan	1.000	19	0.573	0.515	1.000	5	0.510	0.437
Panama	0.791	79	0.738	0.455	1.000	5	0.778	0.595
Paraguay	1.008	4	1.000	0.614	0.769	63	0.768	0.474
Peru	1.000	16	0.905	0.562	0.946	47	0.693	0.532
Philippines	1.000	17	1.000	0.610	1.000	5	0.769	0.534
Poland	0.916	63	0.567	0.433	0.548	87	0.548	0.443
Portugal	1.000	19	0.577	0.468	1.000	5	0.662	0.570
Guinea-Bissau	0.422	117	0.419	0.343				
Qatar	1.000	19	0.727	0.532				
Romania	0.745	87	0.674	0.445	0.652	71	0.652	0.478
Russia	0.654	94	0.592	0.411	0.544	88	0.544	0.415
Saudi Arabia	0.518	112	0.469	0.312	0.758	64	0.408	0.313
Senegal	0.646	95	0.393	0.326	0.608	75	0.395	0.329
Serbia	0.552	107	0.498	0.369	0.569	86	0.567	0.461
Sierra Leone	0.830	74	0.475	0.361	0.577	84	0.572	0.430
Slovakia	0.609	102	0.549	0.395	0.532	90	0.529	0.389
Slovenia	1.000	19	0.617	0.511	1.000	5	0.512	0.493
South Africa	0.561	106	0.529	0.309	0.476	101	0.413	0.289
Zimbabwe	0.762	83	0.410	0.351	0.426	104	0.279	0.235
Spain	1.000	19	0.648	0.546	1.000	5	0.613	0.513
Suriname	0.700	92	0.834	0.524				
Sweden	0.897	65	0.465	0.464	1.000	5	0.460	0.401
Switzerland	1.000	19	1.000	1.000	1.000	5	1.000	0.919
Syria	1.000	19	0.770	0.481				
Thailand	0.871	71	0.717	0.432	1.000	5	0.518	0.384
Togo	1.000	19	0.413	0.346	0.936	48	0.327	0.265
Tunisia	1.000	19	0.599	0.390	0.514	93	0.441	0.333

(Continued)

Table 5.3 (Continued)

	2001–10				2011–20			
Country	Order-*m*	Ranking	FDH	DEA	Order-*m*	Ranking	FDH	DEA
Turkey	0.633	97	0.786	0.565	0.685	69	0.681	0.483
Uganda	0.789	80	0.397	0.343				
Ukraine	0.828	75	0.590	0.415	0.515	92	0.514	0.377
Egypt	0.953	58	0.843	0.526	0.799	59	0.781	0.537
United Kingdom	0.905	64	0.590	0.552	1.000	5	0.722	0.565
Tanzania	1.000	19	0.470	0.424	1.000	5	0.582	0.473
United States	0.979	52	0.752	0.696	0.804	57	0.804	0.668
Burkina Faso	0.752	86	0.350	0.277	0.570	85	0.372	0.288
Uruguay	0.612	99	0.883	0.655	1.000	3	0.703	0.619
Venezuela	1.001	7	0.809	0.483	0.776	61	0.674	0.532

Source: Authors.



Source: Authors.

Figure 5.1 Public sector efficiency 2001–10. Order-m efficiency scores

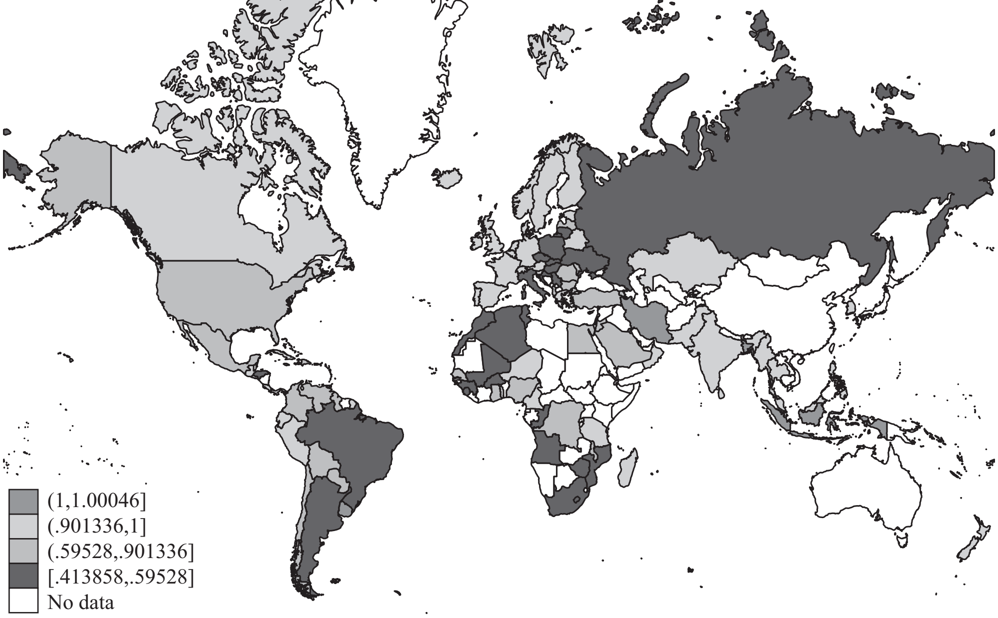


Figure 5.2 Public sector efficiency 2011–20. Order-m efficiency scores

An inspection of the data reveals some first insights. First of all, there is a positive association between the level of development and PSE. More developed countries, at least in Europe and North America, have higher scores. However, higher variation exists in low-income countries, especially in Africa, South America and Asia, where low public sector efficiency

Table 5.4 Correlation of PSE scores

	Correlation of Scores 2001–10		
	Order-*m*	FDH	DEA
Order-*m*	1		
FDH	0.574	1	
DEA	0.715	0.869	1
	Correlation of Scores 2011–20		
	Order-*m*	FDH	DEA
Order-*m*	1		
FDH	0.383	1	
DEA	0.503	0.936	1

Source: Authors.

countries co-exist with countries with a more efficient public sector. Another thing worth noting is the small variation in the efficiency scores within countries over time. This suggests that, even though there are changes in the public sector's production technology, they are somehow common changes between countries.

Comparing the order-*m* scores of the present study with the FDH and DEA scores shows a high degree of correlation. Table 5.4 computes the rank correlation among the three alternative measures to verify this. The table reveals that there is a high-rank correlation among the three methods. The only exception is the correlation between the DEA and order-*m* score for the 2011–20 period. This exercise suggests that all three scores give more or less similar country rankings. Given the advantages of the order-*m* scores, we place higher faith on these rankings.

In the next section, we try to examine more the factors that influence efficiency after reviewing the results of the relevant literature.

3. DETERMINANTS OF PSE

Many studies examine the effect of various determinants on PSE. In section 3.2, we will use the conditional order-*m* efficiency score (Daraio and Simar 2007; De Witte and Kortelainen 2013) to examine the effect of exogenous factors. However, first, we will briefly discuss the economic, institutional, geographical and political factors found to influence PSE in the existing literature. Table 5.5 presents the major determinants of PSE, as already examined in the literature.

3.1 Previous Studies

The variable that has been extensively examined as a determinant of PSE is fiscal decentralization. Theoretically, it appears that there might be two conflicting effects. On the one hand, fiscal decentralization is associated with (i) increased electoral control and (ii) yardstick competition.[7] Thus, politician accountability, and incentives for efficiency, are higher. In contrast, fiscal decentralization negatively affects public good provision technology: in the presence of economies of scale, higher decentralization leads to a higher average cost of production for the public good (Stein 1999). Additionally, national government bureaucracies are more likely to offer talented people better careers and promotion opportunities than local governments, attracting higher-quality individuals (Prud'homme 1995). Finally, other scholars emphasize the potential danger that local politicians and bureaucrats are likely to face, particularly an increase in pressure from local interest groups, with these groups being more influential when the size of the jurisdiction is small (Prud'homme 1995; Bardhan and Mookherjee 2000).

Consequently, when it comes to empirical evidence, the results regarding fiscal decentralization are mixed. For example, Adam et al. (2011b) find a positive effect of fiscal decentralization on PSE. Similarly, Christl et al. (2020), by taking into account endogeneity, find a robust positive impact of decentralization on input-oriented efficiency. However, in contrast to these findings, Adam et al. (2014) show that both opposing forces might co-exist, leading to a non-linear (inverse U-shaped) relationship between PSE and fiscal decentralization. More specifically, their finding is that in relatively centralized systems, an increase in the degree of fiscal decentralization can increase cost due to diseconomies of scale in the provision of public goods, which will be lower than the gains of the electoral accountability, leading to an increase in PSE. In a relatively high decentralization context, however, a further increase in the degree of centralization leads to diseconomies of scale which will prevail over the positive effects of electoral accountability, thus reducing PSE. Similarly, there is large number of studies that examine the relationship at the local level, providing mixed results. For example, Barankay and Lockwood (2007) and Balaguer-Coll et al. (2007) find a positive effect of fiscal decentralization on PSE for Switzerland and Spain, respectively. Still, other studies find a negative association between fiscal decentralization and efficiency (Prud'homme 1995; Stein 1999; Martínez et al. 2018).

Variables related to international trade may also affect PSE. For instance, Hauner and Kyobe (2010) find a negative association between trade openness and educational efficiency, arguing that increased openness will lead to increased exposure to other countries, which will result in technology and skills transfers that increase PSE. Also, Christl et al. (2020) find a positive association between globalization and efficiency, implying spillover effects in adopting good policies.

Population density is an additional technological factor that significantly and positively affects PSE. Higher population density can improve PSE by lowering the cost-of-service provision through economies of scale and lower transportation and heating costs (De Borger and Kerstens 1996a; Afonso and Fernandes 2008; Geys et al. 2010; Hauner and Kyobe 2010; Kalb 2010; Adam et al. 2014). However, in the case of local governments, some studies find a negative association between population density and PSE (for example, Athanassopoulos and Triantis 1998; Kalb et al. 2012; Geys et al. 2013). They argue that in highly concentrated areas, the cost of providing public services is higher due to agglomeration and high complexity, resulting in lower efficiency.

Other population-related characteristics might work similarly to population density. For example, studies examine the age distribution or total population (see Hauner and Kyobe 2010; Christl et al. 2020). The argument regarding age distribution is that a younger population is associated with a higher production cost of education, whereas the older population with a higher cost of the healthcare system. However, when it comes to the total population, the results are mixed since the effects of the economies of scale (for example, De Borger et al. 1994; Bruns and Himmler 2011; Asatryan and De Witte 2015) and scale inefficiencies (Šťastná and Gregor 2015; Geys and Moesen 2008; Ashworth et al. 2014) work in opposite directions.

A simple inspection of Figure 5.1 and Figure 5.2 suggests that income and economic development might determine PSE. However, the results when it comes to formal evidence are somehow mixed. Studies that find a positive effect of income on PSE (Afonso and Fernandes 2008; Afonso et al. 2010; Adam et al. 2011b, 2014; Asatryan and De Witte 2015) highlight the role of high-income elasticity of demand for public sector efficiency. Then again, some studies show that higher income reduces efficiency by increasing the relative cost of public services (Hauner and Kyobe 2010) or, in the case of local governments, by reducing incentives to monitor expenditure when higher incomes create higher local tax revenues (for example, De Borger et al. 1994; De Borger and Kerstens 1996a, 1996b; Bruns and Himmler 2011).

Several other studies point toward the role of the size of the public sector. Even though it is a variable that affects the frontier, it is also considered a factor affecting public sector slack. Hauner and Kyobe (2010) find that public spending reduces public sector efficiency since individuals get declining marginal improvements in performance when spending increases. This result is consistent with several other studies that find a negative correlation between efficiency scores and public spending (Gupta and Verhoeven 2001; Afonso et al. 2005; Herrera and Pang 2005).

When it comes to institutional factors that affect efficiency, the results are limited by the availability of PSE estimates for countries with a low institutional quality. Adam et al. (2011a) show theoretically and empirically that democratic governments tend to be more efficient. They argue that elected officials in democracies are more accountable relative to those in autocracies. Thus, politicians who wish to remain in office should be efficient in producing public goods. Their main finding is that democratic institutions can restrain inefficiencies due to the electoral control mechanism. Asatryan and De Witte (2015) examine whether direct democracy is a determinant of PSE in a sample of German municipalities. Their findings suggest that direct democracy is positively associated with PSE, inducing political competition, thus alleviating principal-agent problems. Also, with voters' direct participation, local institutions of governance become more inclusive, resulting in better economic outcomes. These results are consistent with Adam et al. (2016b) in finding that democratic participation has a positive effect on efficiency (see also Adam et al. 2011a). Similarly, Borge et al. (2008) find a positive impact of democratic participation on PSE for Norway.

Ideology is also found to affect efficiency. Adam et al. (2011b) find that right-wing governments tend to be more efficient. Since left-wing governments support a larger public sector associated with increased production costs, they tend to be more inefficient. Also, Christl et al. (2020) find a negative association between the share of the left and socialist parties in the parliament and efficiency. Similar findings are derived in the literature of local PSE (Borge et al. 2008; Geys et al. 2010; Kalb et al. 2012; Šťastná and Gregor 2015; Helland and Sørensen 2015). However, some find a negative association between ideology and efficiency. For instance, De Borger and Kerstens (1996a, 1996b) find that the presence of socialist parties is positively associated with efficiency.

The structure of the political system can also influence PSE. Adam et al. (2011b) argue that PSE is lower in countries with a majoritarian political system. As the electoral outcome is generally more sensitive to the incumbent's performance in majoritarian-type elections, elected officials have strong incentives to perform well, stimulating public sector performance. Finally, the authors also find that a higher number of spending ministers lowers PSE. This result is consistent with the idea that diseconomies of scale may be present within the government, leading to diminished government output (see Mierau et al. 2007).

Likewise, some studies examine the effect of political concentration as a proxy for political strength. The results, however, are again mixed: when there is high political concentration, the imposition of fiscal constraints is easier, resulting in higher efficiency (for example, Borge et al. 2008; Bruns and Himmler 2011; Afonso and Aubyn 2013; Šťastná and Gregor 2015). Conversely, expenditure monitoring by the opposition is limited, resulting in lower efficiency (for example, Balaguer-Coll et al. 2007; Geys et al. 2010; Kalb 2010; Helland and Sørensen 2015). Similarly, Adam et al. (2011a) also find that coalition governments are negatively correlated with efficiency. The rationale behind this result is that the likelihood of a favorable to efficiency policy change decreases as the number of 'veto-players' increases. As coalition governments are characterized by a larger number of veto players than one-party majority governments, they have a stronger 'status quo bias' for established policies, even when they are inefficient. The same result is obtained in De Borger et al. (1994), which examines the case of Belgian municipalities. The durability of the government as a proxy of volatility is also found to affect efficiency (Hauner and Kyobe 2010) positively.

Table 5.5 Determinants of PSE

Variable	Definition	Source	Expected Effect	Related Literature
Executive Ideology	Ordinal variable, 0=not defined, 1=right, 2=center, 3=left	Database of Political Institutions	Left less efficient	Christl et al. (2020), Helland and Sørensen (2015), Adam et al. (2011b), Geys et al. (2010), De Borger and Kerstens (1996a,b)
Electoral system	Dummy =1 if parliamentary system, 0 otherwise	Database of Political Institutions	Favorable	Christl et al. (2020), Adam et al. (2014),
Democracy	Combined Freedom House and Polity IV country score, higher values more democratic ideals	QoG- Freedom House and Polity Project	Favorable	Asatryan and De Witte (2015), Bosch-Roca et al. (2012), Geys et al. (2010)
Voter Turnout	Percentage of total population that voted in the previous elections	QoG- Vanhanen (2019)	Ambiguous	Šťastná and Gregor (2015), Asatryan and De Witte (2015), Giordano and Tommasino (2013), Geys et al. (2010).
Population Density (log)	People per sq. km of land area	QoG- WDI	Unfavorable	Adam et al. (2014), Geys et al. (2010), Hauner and Kyobe (2010)
GDP per capita (log)	Log of GDP per capita, 2011 constant Purchasing Power Parity values	QoG- WDI	Ambiguous	Afonso et al. (2010), Hauner and Kyobe (2010), Geys et al. (2010), De Borger et al. (1994)
Openness	Imports plus exports over GDP	QoG- WDI	Ambiguous	Christl et al. (2020), Hauner and Kyobe (2010)
Government Fractionalization	Herfindahl Index of Fractionalization in the government gives the probability that two random draws will produce legislators from different parties	Database of Political Institutions	Unfavorable	Sørensen (2014), Geys et al. (2010), De Borger et al. (1994), Eeckaut et al. (1993)

(Continued)

Table 5.5 *(Continued)*

Variable	Definition	Source	Expected Effect	Related Literature
Fiscal Decentralization	Synthetic Index of expenditure decentralization, dependence on grants, taxation autonomy and borrowing freedom	Maksym and Sha (2014)	Inverted U	Adam et al. (2014), Adam et al. (2011b), Barankay and Lockwood (2007), Balaguer-Coll et al. (2007)
Population(log)	Log of total population	QoG- WDI	Ambiguous	Christl et al. (2020), Asatryan and De Witte (2015), Hauner and Kyobe (2010), De Borger et al. (1994)
Government Spending	Government Consumption as a share of GDP	QoG- WDI	Unfavorable	Hauner and Kyobe (2010), Afonso et al. (2010), Gupta and Verhoeven (2001)
2011–2020 dummy	Dummy =1 if the PSE score is for 2011–20		Ambiguous	
Regional Dummies	Regions according to UN Stats Regions		Ambiguous	

Source: Authors.

3.2 Some Evidence

In this section, we use the efficiency scores as computed in section 2.2 to examine the validity of previous findings in the literature regarding the effect of external factors on PSE in a unified framework. Typically, the literature has employed OLS regression and censored variables analysis for this task. However, Simar and Wilson (2007, 2011) criticize this approach. First, the standard OLS misses that PSE scores are bounded in the *[0,1]* range. Moreover, even in the censored variables framework, the standard regression analysis does not consider the correlation of the PSE scores; all efficiency scores are computed from the same sample, that is, the relative efficiency score for each country depends on the efficiency score of the rest. Finally, the standard DEA or FDH scores are computed with the efficient score (that is, *1* in the case of input-oriented efficiency scores) being the target value. For these reasons, Simar and Wilson (2007) propose a double bootstrap technique to determine the effect of exogenous factors on efficiency scores.

However, even this bootstrap technique has several disadvantages. First, it only examines deviations from the frontier. Yet, external factors might be equally responsible for the formation and the shift of the frontier. For example, the discussion in the previous section has suggested that factors that shape the frontier also affect the distance to the frontier (for example, size of government). Moreover, the second stage regression relies on some parametric assumptions (for example, linear model and truncated normal error term).

To overcome these limitations, we follow Daraio and Simar (2007) and calculate the conditional order-*m* scores, that is, input efficiency scores as in section 2.2 but conditional on the exogenous factors.[8] We, then, take the share of the conditional efficiency scores on the unconditional efficiency scores and estimate a smooth non-parametric kernel regression line on the external factor as in De Witte and Kortelainen (2013). For the case of the multivariate analysis, we use partial regression plots to visualize the effect and a non-parametric bootstrap procedure to obtain standard errors and *p-values* for the influence of each external factor on the ratio of efficiency (Li and Racine 2007; De Witte and Kortelainen 2013; Asatryan and De Witte 2015). The visualization of the partial plots determines the average effect we are interested in here.

Based on the discussion in the previous subsection, we examine the effect of various variables, as also examined in the existing literature, on PSE. The following table presents these variables, their sources, the paper where they were introduced, and their expected effect on PSE. As most studies were about OECD countries, some variables were not available in our sample, or the data source is different. In Table 5.6, we present the results of our analysis. In Figure 5.3, we present the partial regression plots, as in the third column, that is, the model with only the statistically significant variables.[9]

As we have input-oriented efficiency scores, an unfavorable external variable acts as an extra input in the production process requiring more inputs to produce an additional unit of output. In this case, the conditional PSE score will be higher than the unconditional PSE for larger external variable values. In contrast, a favorable external variable acts as a substitutive input, requiring fewer inputs to produce an extra unit of output, resulting in a lower conditional PSE score. Hence, a positively sloped regression line indicates a negative effect on PSE, whereas a negatively sloped line indicates a positive effect (Daraio and Simar 2007).

The table shows that most of the variables turn out as significant determinants of PSE. Moreover, their effect appears consistent across all specifications. Interestingly, in most cases, the indicated influence found here verifies the findings of the existing literature. The

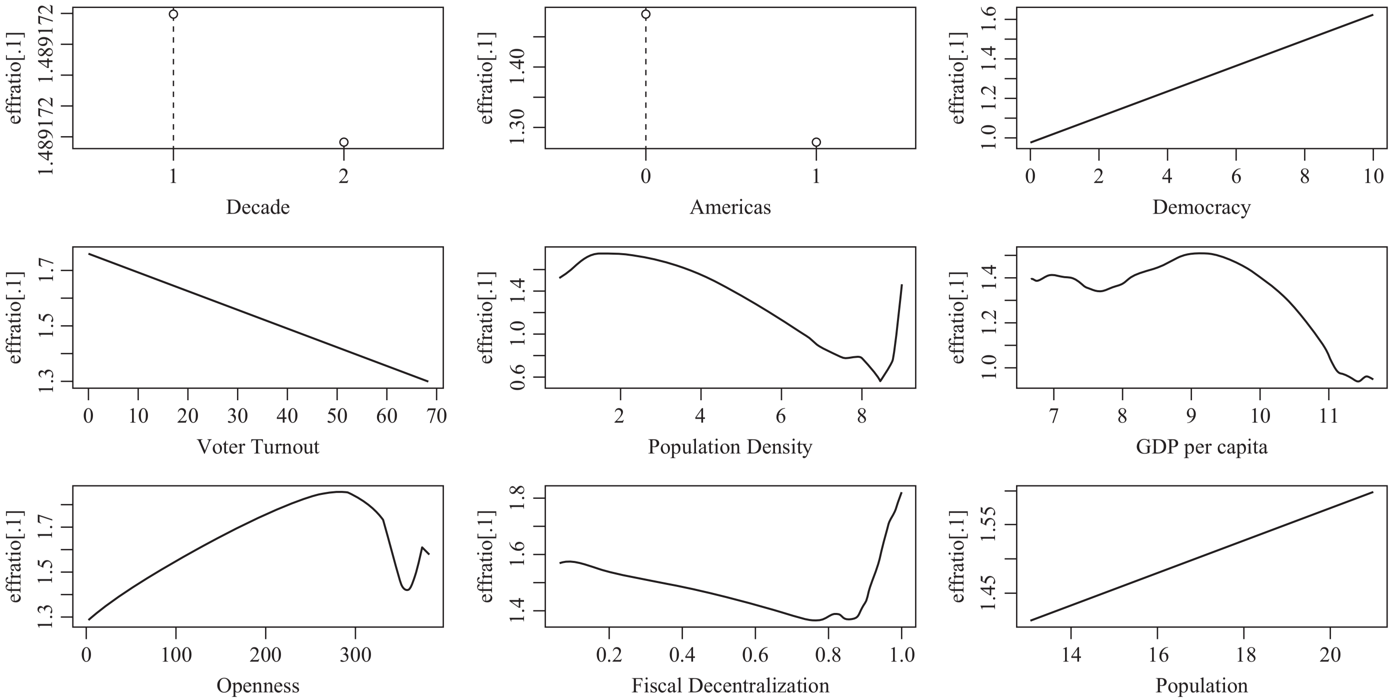
Table 5.6 Results

Variable	Unconditional	Baseline Model		Only Statistically Significant		Including Spending		No Regional Dummies		Only Democracy	
		Influence	*p*-value	Influence	*p*-value	Influence	*p*-value	Influence	*p*-value	Influence	*p*-value
Executive Ideology		Center more efficient				Left less efficient	**	Center more efficient			
Electoral system		varying				varying		varying			
2011–20		favorable	*	favorable		favorable		favorable		favorable	
Africa		unfavorable				unfavorable				unfavorable	
Americas		favorable	*	favorable		favorable				favorable	***
Asia		unfavorable				favorable				favorable	*
Europe		favorable				favorable				unfavorable	**
Oceania		unfavorable				varying				favorable	
Democracy		unfavorable	***	unfavorable	***	unfavorable	***	unfavorable	***	unfavorable	***
Voter Turnout		favorable	***	favorable	***	favorable		favorable	***		
Population Density		favorable	***	favorable	***	varying	***	favorable	***	favorable	***
GDP per capita (log)		favorable	***	favorable	***	favorable	***	favorable	*	favorable	***
Openness		varying	***	varying	***	varying	***	varying	***	varying	

Table 5.6 (Continued)

Variable	Unconditional	Baseline Model		Only Statistically Significant		Including Spending		No Regional Dummies Dummies		Only Democracy	
		Influence	*p*-value	Influence	*p*-value	Influence	*p*-value	Influence	*p*-value	Influence	*p*-value
Government Fractionalization		varying				varying	***	varying			
Fiscal		inverted-U	***	inverted-U	***	unfavorable	**	varying		inverted-U	***
Population(log)		varying	**	unfavorable	***	unfavorable		varying		unfavorable	**
Government Spending						unfavorable	***				
Mean efficiency score	0.794										
Standard deviation of mean efficiency score	0.201										
Observations	213										
R2		0.01		0.53		0.01		0.04		0.61	

Notes: Table presents estimates of the share of order-*m* unconditional to conditional efficiency, on each variable. A favorable association indicates an efficiency-enhancing effect of the variable. No effect indicates that at the median we could not find either a positive or negative association. ***, **, * indicate *p*-value less than 10%, 5%, 1%, respectively. Besides the baseline model, we also estimate the same model including only the variables found to exert statistically significant influence in the baseline model, when we include the government spending as a potential determinant of efficiency, when we exclude the regional dummies, and finally when we exclude all political variables besides democracy from the model.



Source: Authors.

Figure 5.3 Partial regression plots

only notable exception to this is *Democracy*, which appears to exert a significant unfavorable influence in all columns, in sharp contrast to Adam et al. (2011a). We should also note that *Fiscal Decentralization*, which has been extensively examined in the literature, has an inverted U-shaped relationship with PSE. This indicates an optimal degree of decentralization, as found in Adam et al. (2014).

Concerning the rest of the variables, as expected, richer countries tend to have higher PSE. Also, as expected, when we introduce the size of the government sector, we find that a larger public sector tends to be more inefficient. Similarly, higher population density for most of the values is associated with higher PSE, indicating that economies of scale are important. Interestingly, the partial regression also verifies that after a value of *Population Density*, diseconomies of scale start to kick in, inducing higher inefficiency. Regarding the total population, the influence appears to be unfavorable, suggesting that larger countries exhibit higher public sector inefficiency. Of course, this is consistent with the idea that there are economies of scale in the production of the public good. Finally, *Openness* has a varying effect. Thus, we cannot determine the direction of its influence.

For the rest of the political variables, besides democracy, voter turnout is associated with higher PSE. On the other hand, ideology does not appear to have a robust effect, even though, in all cases, chief executives at the political center appear to be more efficient in providing public goods.

Finally, when it comes to regional dummies, the robust finding is that in the Americas, PSE is higher. At the same time, there appears that in the 2011–20 period, overall PSE has increased.

4. CONCLUSIONS

In this chapter, we have reviewed the literature that examines the measurement and determinants of overall PSE. As we have highlighted several disadvantages in the existing methods employed in the literature, we have used Partial Efficiency Analysis, and specifically order-*m* efficiency scores to construct measures of PSE, for a sample of 114 countries, over the 2000–20 period. We have then used these scores to examine the effect of external factors on PSE. Overall, our findings show that poorer countries with a larger public sector tend to be more inefficient. Moreover, we have found that political institutions are crucial determinants of PSE. Finally, our results show that economies of scale in the production of the public good significantly affect relative efficiency scores.

NOTES

1. Of course, parametric methods do not face such problems, but have the disadvantage that an underlying functional form must be imposed.
2. Regarding local governments, the interested reader is referred to the recent and extensive reviews of Narbón-Perpiñá and De Witte (2018a, 2018b).
3. Namely, the analysis (i) does not impose a specific functional form on the underlying relationship between external variables and PSE and (ii) allows the researcher to evaluate the effect of both variables that determine the shape of the frontier as well as the distance from the frontier (Daraio and Simar 2005).
4. This latter criticism does not apply to parametric approaches, which take into account the outliers by using a statistical model, however they impose a specific functional form or distribution for the errors (for example stochastic frontier and so on).

5. According to Daraio and Simar (2007), a reasonable choice of D is 200.
6. Note that the FDH and the order-*m* give the same efficiency scores when *m* approaches infinity. The order-*m* estimations in this and the following section were made in R 4.0.2 software (R Core Team, 2019) using the np package (Hayfield and Racine 2008). We would like to thank professor De Witte for giving us access to his R code.
7. That is, citizens are in advantage when they can evaluate their policymakers' performance by comparing it with the performance of their neighbors, (see also Besley and Case 1995)
8. We should note that the (unconditioned) FDH input-oriented measure conditions on the amount of the output being equal or less than the boundary of the FDH of the sample. In contrast, the conditional FDH conditions on the amount of the output being equal to or less than the boundary of the FDH of the sample and on the amount of the external variable being **equal** to the value of the external factor on the boundary. Then, the estimation of the conditional order-*m* efficiency requires the estimation a non-parametric smoothed estimator. We use a Epanechnikov kernel function and the Least Squares Cross Validation (Bădin et al. 2010) to obtain the bandwidth.
9. In the results that follow we have estimated the order-*m* efficiency scores for the whole set of countries for both decades (2001–10 and 2011–20). Then, the time effect is estimated as an external factor in the conditional PSE scores.

REFERENCES

Adam, A., M. D. Delis and P. Kammas (2011a), 'Are democratic governments more efficient?', *European Journal of Political Economy*, **27** (1), 75–86.

Adam, A., M. Delis and P. Kammas (2011b), 'Public sector efficiency: leveling the playing field between OECD countries', *Public Choice*, **146** (1), 163–83.

Adam, A., M. D. Delis and P. Kammas (2014), 'Fiscal decentralization and public sector efficiency: evidence from OECD countries', *Economics of Governance*, **15** (1), 17–49.

Afonso, A. and M. S. Aubyn (2013), 'Public and private inputs in aggregate production and growth: a cross-country efficiency approach', *Applied Economics*, **45** (32), 4487–502.

Afonso, A. and S. Fernandes (2008), 'Assessing and explaining the relative efficiency of local government', *The Journal of Socio-Economics*, **37** (5), 1946–79.

Afonso, A., L. Schuknecht and V. Tanzi (2005), 'Public sector efficiency: an international comparison', *Public Choice*, **123** (3), 321–47.

Afonso, A., L. Schuknecht and V. Tanzi (2010), 'Public sector efficiency: evidence for new EU member states and emerging markets', *Applied Economics*, **42** (17), 2147–64.

Angelopoulos, K., A. Philippopoulos and E. Tsionas (2008), 'Does public sector efficiency matter? Revisiting the relation between fiscal size and economic growth in a world sample', *Public Choice*, **137** (1), 245–78.

Asatryan, Z. and K. De Witte (2015), 'Direct democracy and local government efficiency', *European Journal of Political Economy*, **39**, 58–66.

Ashworth, J., B. Geys, B. Heyndels and F. Wille (2014), 'Competition in the political arena and local government performance', *Applied Economics*, **46** (19), 2264–76.

Athanassopoulos, A. D. and K. P. Triantis (1998), 'Assessing aggregate cost efficiency and the related policy implications for Greek local municipalities', *INFOR: Information Systems and Operational Research*, **36** (3), 66–83.

Bădin, L., C. Daraio and L. Simar (2010), 'Optimal bandwidth selection for conditional efficiency measures: a data-driven approach', *European Journal of Operational Research*, **201** (2), 633–40.

Balaguer-Coll, M. T., D. Prior and E. Tortosa-Ausina (2007), 'On the determinants of local government performance: A two-stage non-parametric approach', *European Economic Review*, **51** (2), 425–51.

Barankay, I. and B. Lockwood (2007), Decentralization and the productive efficiency of government: evidence from Swiss cantons', *Journal of Public Economics*, **91** (5–6), 1197–218.

Bardhan, P. K. and D. Mookherjee (2000), 'Capture and governance at local and national levels', *American Economic Review*, **90** (2), 135–9.

Besley, T. and A. Case (1995), 'Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition', *American Economic Review*, *90* (1), 25–45.

Borge, L.-E., T. Falch and P. Tovmo (2008), 'Public sector efficiency: the roles of political and budgetary institutions, fiscal capacity, and democratic participation', *Public Choice*, **136** (3), 475–95.

Bruns, C. and O. Himmler (2011), 'Newspaper circulation and local government efficiency', *The Scandinavian Journal of Economics*, **113** (2), 470–92.

Cazals, C., J.-P. Florens and L. Simar (2002), 'Non-parametric frontier estimation: a robust approach', *Journal of Econometrics*, **106** (1), 1–25.

Christl, M., M. Köppl-Turyna and D. Kucsera (2020), 'Determinants of public-sector efficiency: decentralization and fiscal rules', *Kyklos*, **73** (2), 253–90.

Daraio, C. and L. Simar (2005), 'Introducing environmental variables in non-parametric frontier models: a probabilistic approach', *Journal of Productivity Analysis*, **24** (1), 93–121.
Daraio, C. and L. Simar (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*, Springer Science & Business Media.
De Borger, B. and K. Kerstens (1996a), 'Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches', *Regional Science and Urban Economics*, **26** (2), 145–70.
De Borger, B. and K. Kerstens (1996b), 'Radial and nonradial measures of technical efficiency: an empirical illustration for Belgian local governments using an FDH reference technology', *Journal of Productivity Analysis*, **7** (1), 41–62.
De Borger, B., K. Kerstens, W. Moesen and J. Vanneste (1994), 'Explaining differences in productive efficiency: an application to Belgian municipalities', *Public Choice*, **80** (3), 339–58.
De Witte, K. and M. Kortelainen (2013), 'What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables', *Applied Economics*, **45** (17), 2401–12.
Vanden Eeckaut, P., H. Tulkens and M.-A. Jamar (1993), 'Cost efficiency in Belgian municipalities', in Harold O. Fried, C.A. Knox Lovell and Shelton S. Schmidt (eds). *The Measurement of Productive Efficiency. Techniques and Applications*, Oxford University Press, New York, Oxford, 300–334.
Geys, B., F. Heinemann and A. Kalb (2010), 'Voter involvement, fiscal autonomy and public sector efficiency: evidence from German municipalities', *European Journal of Political Economy*, **26** (2), 265–78.
Geys, B., F. Heinemann and A. Kalb (2013), 'Local government efficiency in German municipalities', *Raumforschung Und Raumordnung*, **71** (4), 283–93.
Geys, B. and W. Moesen (2008), *Exploring Sources of Local Government Technical Inefficiency: Evidence from Flemish Municipalities*, Discussion Papers, Research Unit: Market Processes and Governance SP II 2008–18, WZB Berlin Social Science Center.
Giordano, R. and P. Tommasino (2013), 'Public-Sector Efficiency and Political Culture', *FinanzArchiv: Public Finance Analysis*, *69* (3), 289–316.
Gupta, S. and M. Verhoeven (2001), 'The efficiency of government expenditure: experiences from Africa', *Journal of Policy Modeling*, **23** (4), 433–67.
Hauner, D. and A. Kyobe (2010), 'Determinants of government efficiency', *World Development*, **38** (11), 1527–42.
Hayfield, T. and J. S. Racine (2008), 'Nonparametric econometrics: the np package', *Journal of Statistical Software*, **27** (5), 1–32.
Helland, L. and R. J. Sørensen (2015), 'Partisan bias, electoral volatility, and government efficiency', *Electoral Studies*, **39**, 117–28.
Herrera, S. and G. Pang (2005), *Efficiency of Public Spending in Developing Countries : An Efficiency Frontier Approach*, Policy Research Working Paper Series 3645, The World Bank, June.
Kalb, A. (2010), 'The impact of intergovernmental grants on cost efficiency: theory and evidence from German municipalities', *Economic Analysis and Policy*, **40** (1), 23–48.
Kalb, A., B. Geys and F. Heinemann (2012), 'Value for money? German local government efficiency in a comparative perspective', *Applied Economics*, **44** (2), 201–18.
Li, Q. and J. S. Racine (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton N.J.: Princeton University Press.
Martínez, Y. U., P. P. Arzoz and B. I. Apezteguía (2018), 'Does decentralization contribute to efficiency? Evidence from OECD countries', *Applied Economics*, **50** (7), 726–42.
Ivanyna, M. and A. Shah (2014), 'How Close Is Your Government to Its People? Worldwide Indicators on Localization and Decentralization', *Economics: The Open-Access, Open-Assessment E-Journal*, 8 (2014–3), 1.
Mierau, J. O., R. Jong-A-Pin and J. de Haan (2007), 'Do political variables affect fiscal policy adjustment decisions? New empirical evidence', *Public Choice*, **133** (3), 297–319.
Narbón-Perpiñá, I. and K. De Witte (2018a), 'Local governments' efficiency: a systematic literature review—part I', *International Transactions in Operational Research*, **25** (2), 431–468.
Narbón-Perpiñá, I. and K. De Witte (2018b), 'Local governments' efficiency: a systematic literature review—part II', *International Transactions in Operational Research*, **25** (4), 1107–1136.
Papaioannou, S. (2016), 'Public investment multipliers in EU countries: Does the efficiency of public sector matter?', *MPRA Paper 70332, University Library of Munich*, Germany.
Park, B. U., L. Simar and Ch. Weiner (2000), 'The FDH estimator for productivity efficiency scores: asymptotic properties', *Econometric Theory*, **16** (6), 855–77.
Prud'homme, R. (1995), 'The dangers of decentralization', *The World Bank Research Observer*, **10** (2), 201–20.
R Core Team (2019), *Core. R: A Language and Environment for Statistical Computing. 2013*.
Simar, L. and P. W. Wilson (2007), 'Estimation and inference in two-stage, semi-parametric models of production processes', *Journal of Econometrics*, **136** (1), 31–64.
Simar, L. and P. W. Wilson (2011), 'Two-stage DEA: caveat emptor', *Journal of Productivity Analysis*, **36** (2), 205–18.
Sørensen, R. J. (2014), 'Political competition, party polarization, and government performance', *Public Choice, 161* (3), 427–50.

Šťastná, L. and M. Gregor (2015), 'Public sector efficiency in transition and beyond: evidence from Czech local governments', *Applied Economics*, **47** (7), 680–99.
Stein, E. (1999), 'Fiscal decentralization and government size in Latin America', *Journal of Applied Economics*, **2** (2), 357–91.
Tauchmann, H. (2012), 'Partial frontier efficiency analysis', *The Stata Journal: Promoting Communications on Statistics and Stata*, **12** (3), 461–78.
Teorell, J., S. Dahlberg, S. Holmberg, B. Rothstein, N. Alvarado Pachon and S. Axelsson (2020), *The Quality of Government Standard Dataset, Version Jan20*, University of Gothenburg: The Quality of Government Institute, accessed 29 September 2020 at http://www.qog.pol.gu.se doi:10.18157/qogstdjan20.

6. On the persistence of public sector efficiency and the role of historical institutional quality

Konstantinos Angelopoulos and Pantelis Kammas

1. INTRODUCTION

The role and the size of the public sector have increased since the beginning of the twentieth century, and especially since the 1960s. Nowadays, the state intervenes in almost all areas of economic activity and also absorbs around 50 per cent of GDP in many advanced economies. It is thus natural that its performance and efficiency are considered to be of key importance to the functioning and prospects of the society. Given the scarcity of social resources, understanding the factors that shape public sector efficiency is important to improve socioeconomic outcomes. In a series of papers, Afonso *et al.* (2005, 2010, 2020), Afonso and St Aubyn (2005, 2006) and Afonso and Kazemi (2016) have constructed measures of public sector performance and public sector efficiency for a number of countries for recent decades. These measures have highlighted differences in public sector performance and efficiency across countries, and indicate areas where policy interventions and reforms are more needed.

In this paper, we aim to investigate the long-run persistence of comparative public sector efficiency, across OECD countries, and the importance of institutional fundamentals in shaping it. We focus on core institutional fundamentals, which have been widely acknowledged to impact economic outcomes.[1] Core institutional fundamentals may influence the performance of the public sector. They can also lead to rent-seeking activities that imply misallocation of public funds, leading to spending slack and misallocation of public sector employee effort. Because the effect of institutional fundamentals on economic outcomes such as public sector efficiency can be persistent, we investigate public sector efficiency in conjunction with such features over a long time period.

Our first objective is to provide a first attempt to examine the evolution of public sector efficiency across countries over a long time period. Existing research on public sector efficiency has focused mainly on more recent decades, given data availability and the natural motivation to aid policy making in modern economies. To complement this body of work, we begin our analysis by constructing a consistent proxy of the efficiency of the public sector, for what is now known as a group of developed OECD countries, since the 1900. We use public spending data from Mauro *et al.* (2015) on primary government expenditure.[2] Naturally, measuring the performance of the public sector across 21 countries, over a century, using consistent measures, requires that outputs that are affected by the public sector are approximated by key variables that are well measured over time and countries, and that have been relevant as objectives for the public sector since 1900. We thus focus on measures of education and health outcomes.[3]

Our first finding is that, although for many countries public sector efficiency, understood as the ability to transform inputs into outputs relative to the frontier, has improved, comparative public sector efficiency shows persistence. In particular, countries with a relatively (in) efficient public sector back in the 1900s, at the beginning of the significant rises in government spending in modern nation states, tend to also have a relatively (in)efficient public sector in more recent decades. This confirms the view that the shaping of good public sector efficiency takes a long time to develop and that initial conditions or events from even the distant past can have long-lasting effects.

Our next objective is to examine the role of institutional quality in generating persistence of public sector efficiency. We present a stylised model to study the channels via which the historical experience of institutional quality in a country can affect future public sector efficiency. In particular, we focus on two channels regarding the effect of pre-determined institutions: on public sector output (performance), and on rent-seeking activities, which imply misallocation of public funds and of government officials' effort from producing the public sector output. We examine the relationship between past and present institutional quality and public sector efficiency in an environment where institutional quality shapes the behaviour of government officials, and in turn the efficiency of the public sector, whilst being determined by the very behaviour it has influenced over time.

We investigate the empirical relationship between proxies of institutional quality, using data from Coppedge *et al.* (2019), and the measures of public sector efficiency we have constructed. We find that there is a positive relationship between early institutional quality, at the turn of the twentieth century, and public sector efficiency in the second half of the twentieth century in education. Public sector efficiency in health in the second half of the twentieth century has a weaker relationship with institutional quality in the period around 1900. We study further the link between institutional quality and public sector efficiency by focusing on the channel of public sector performance. We find, for both education and health, a strong relationship between public sector performance in the second half of the twentieth century and institutional quality at the beginning of the century. Taken together, our findings suggest that the dependence of public sector performance on past institutions is an important factor in understanding the persistence of public sector efficiency.

2. PUBLIC SECTOR EFFICIENCY OVER THE TWENTIETH CENTURY

We construct alternative sub-indices of relative Public Sector Efficiency (PSE) by comparing public expenditure to output in two areas that have formed objectives for government involvement over the twentieth century for a set of 21 OECD countries. In particular, we construct PSE measures over the five decades between the 1950s and the 2000s, and also for the decade of the 1900s. More precisely, we seek to develop PSE scores when public sector output is approximated by measures of Education and Health, by using variables that are the same in spirit to those employed by Afonso *et al.* (2005) and by employing a simple input-oriented, decreasing returns to scale Data Envelopment Analysis (DEA) methodology.

The long time dimension of our analysis (that extends back to the early twentieth century) constrains data availability concerning government spending in specific policy areas. To create a comparable dataset across time and countries, we proceed by employing as Public Sector Expenditure (PEX) the primary expenditure data from the historical fiscal database developed by Mauro *et al.* (2015). The database developed by Mauro *et al.* (2015) draws on historical records for 55 countries over the past two centuries and to the best of our knowledge is one the most reliable and comprehensive databases available on fiscal flows and stocks. Primary expenditure (as a per cent of GDP)[4] is not an ideal variable to investigate public sector efficiency in specific areas, given that it is general spending. However, it has the important advantage that it has been constructed based on a common classification and therefore is suitable for comparisons between countries which is a necessary condition for our analysis.[5]

To estimate PSE, we require measures of Public Sector Performance (PSP) that relate to outputs of public spending. We follow the logic in the relevant literature (see for example Afonso *et al.* 2005) and adjust this to data availability in historical periods. In particular, we first consider education as a target for government involvement, and as a measure we employ the average years of total schooling in total population, developed by Lee and Lee (2016) (defined as the average years of schooling among the adult population as a whole and at all the levels of education). We then consider health as a target for government involvement, and to measure health outcomes we employ the inverse of the infant mortality rate (per 1000 live births). Data for infant mortality (per 1000 live births) are obtained from the database of the World Health Organization for the 1950s and onwards and from Mitchell (2013) for the decade of 1900s. Details of data and measures we use are in Table 6.1.

We focus on productive efficiency, which reflects the ability of a public sector to maximise output for a given set of inputs. Countries that succeed are on the frontier, while those that do not optimise the use of their inputs are inefficient to a variable extent. This type of inefficiency can be measured in terms of the distance of a given country from the best-practice equivalent (which forms the frontier) and the resulting score is a scalar measure ranging between zero (the lowest efficiency score) and one (the best-practice public sector). We employ simple input-oriented, decreasing returns to scale DEA to obtain evaluations of PSE.[6] Therefore, each score obtained reflects the distance between the respective data point (country) and the best practice

Table 6.1 PSE inputs (PEX) and outputs (PSP)

Public Sector Efficiency (PSE) in the policy area of *Education*		
Inputs (PEX) and Outputs (PSP)	Description	Sources
Public Expenditure (PEX)	Primary government expenditure. Total government expenditure minus interest paid on public debt (per cent of GDP).	Mauro et al. (2015)
Public Sector Performance (PSP)	Average years of total schooling in total population.	Lee and Lee (2016)
Public Sector Efficiency (PSE) in the policy area of *Health*		
Inputs (PEX) and Outputs (PSP)	Description	Sources
Public Expenditure (PEX)	Primary government expenditure. Total government expenditure minus interest paid on public debt (per cent of GDP).	Mauro et al. (2015)
Public Sector Performance (PSP)	Inverse of infant mortality (per 1000 births) (that is, 1000-Infant mortality (per 1000 births)).	Own calculations based on World Health Organization (2020) *World Health Data Platform* and Mitchell (2013)

country that shapes the frontier. The estimation methodology described above is applied on a set of 21 OECD countries over five decades, from the 1950s to 2000s but also for the decade of the 1900s.[7] These countries are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, New Zealand, Norway, the Netherlands, Portugal, Spain, Sweden, Switzerland, the United Kingdom and the United States.

In Tables 6.2 and 6.3 we present the PSE scores obtained working as described above. PSE indices using Education or Health as proxies of government performance improve over time for many countries. This means that given amounts of government spending achieve better Public Sector Outputs (Performance) (that is higher levels of schooling, lower levels of infant mortality) relative to the efficiency frontier. However, in terms of relative positions with respect to PSE, there is persistence, which extends back to the 1900s, at the beginning of the significant rises in government spending in modern nation states. Therefore, countries

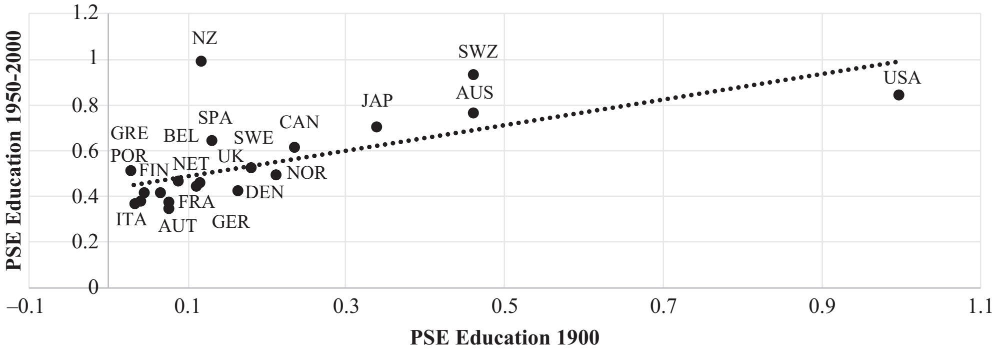
Table 6.2 Public sector efficiency (PSE) in the policy area of Education

Countries	PSE 1900	PSE 1950	PSE 1960	PSE 1970	PSE 1980	PSE 1990	PSE 2000
Australia	0.46	0.55	0.36	0.88	0.96	0.91	0.84
Austria	0.08	0.17	0.14	0.38	0.41	0.45	0.49
Belgium	0.09	0.27	0.30	0.42	0.50	0.63	0.59
Canada	0.24	0.46	0.28	0.65	0.70	0.73	0.82
Denmark	0.17	0.26	0.26	0.41	0.49	0.55	0.55
Finland	0.04	0.15	0.18	0.46	0.48	0.43	0.52
France	0.08	0.16	0.19	0.49	0.37	0.47	0.51
Germany	0.07	0.20	0.24	0.36	0.44	0.55	0.63
Greece	0.03	0.18	0.29	0.59	0.66	0.71	0.61
Ireland		0.22	0.19	0.46	0.59	0.80	0.80
Italy	0.05	0.22	0.23	0.38	0.47	0.54	0.57
Japan	0.34	0.31	0.35	0.76	0.92	0.98	0.87
The Netherlands	0.12	0.23	0.25	0.45	0.50	0.61	0.62
New Zealand	0.12	1.00	1.00	1.00	0.92	0.97	1.00
Norway	0.22	0.36	0.23	0.47	0.57	0.60	0.68
Portugal	0.04	0.17	0.16	0.30	0.49	0.52	0.48
Spain	0.13	0.36	0.34	1.00	0.79	0.63	0.66
Sweden	0.18	0.27	0.28	0.66	0.75	0.57	0.55
Switzerland	0.46	1.00	1.00	0.89	1.00	0.85	0.77
United Kingdom	0.11	0.18	0.19	0.45	0.54	0.64	0.63
United States	1.00	0.46	0.69	0.86	1.00	1.00	1.00

Table 6.3 Public sector efficiency (PSE) in the policy area of Health

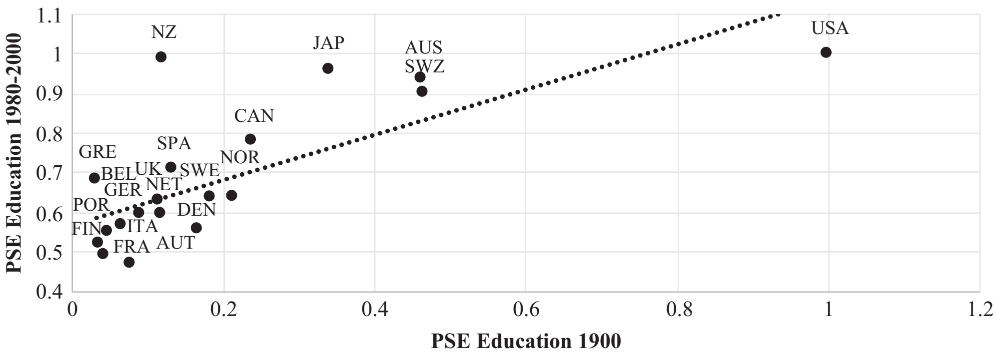
Countries	PSE 1900	PSE 1950	PSE 1960	PSE 1970	PSE 1980	PSE 1990	PSE 2000
Australia		1.00	0.37	0.69	0.77	0.89	0.91
Austria	0.51	0.26	0.21	0.28	0.48	0.58	0.65
Belgium	0.44	0.34	0.37	0.37	0.50	0.69	0.69
Canada		0.51	0.30	0.55	0.66	0.75	0.84
Denmark	0.92	0.53	0.55	0.60	0.53	0.57	0.61
Finland		0.33	0.58	0.84	0.64	0.54	0.69
France	0.46	0.33	0.39	0.89	0.52	0.59	0.63
Germany	0.22	0.69	0.30	0.32	0.54	0.66	0.73
Greece	0.64	0.31	0.36	0.46	0.76	0.84	0.75
Ireland		0.29	0.24	0.39	0.55	0.84	0.88
Italy	0.41	0.42	0.40	0.35	0.59	0.70	0.74
Japan		0.38	0.42	1.00	1.00	1.00	1.00
The Netherlands	0.47	0.79	0.95	0.55	0.50	0.63	0.70
New Zealand		0.42	0.37	0.67	0.69	0.93	1.00
Norway	1.00	0.86	0.58	0.60	0.58	0.61	0.77
Portugal		0.68	0.56	0.45	0.78	0.82	0.76
Spain	0.71	0.82	0.70	1.00	1.00	0.87	0.84
Sweden	1.00	1.00	1.00	1.00	0.75	0.57	0.64
Switzerland	1.00	1.00	1.00	0.88	0.92	0.91	0.93
United Kingdom	0.53	0.25	0.25	0.43	0.57	0.76	0.72
United States		0.46	0.28	0.52	0.71	0.88	0.85

that find themselves in a relatively worse position at the turn of the twentieth century, tend to have lower PSE scores even in the more recent decades and vice versa. Specifically, as can be seen in Figures 6.1–6.4, there is a positive association between the PSE score that a country achieves during the 1900s and the corresponding PSE score of the same country during the 1950–2010 period or the 1980–2010 period. Countries that were relatively efficient during the early twentieth century (like for example United States, Switzerland and Australia in the case of Education or United States, Japan and Canada in the case of Health) remained in a relatively efficient position also during the more recent decades. In contrast, countries that were relatively inefficient during the early twentieth century (like for example Greece, Portugal and Spain) appear to share relatively lower contemporaneous PSE scores. This finding suggests that initial conditions or events from even the distant past can have long-lasting effects on the efficiency of the public sector.



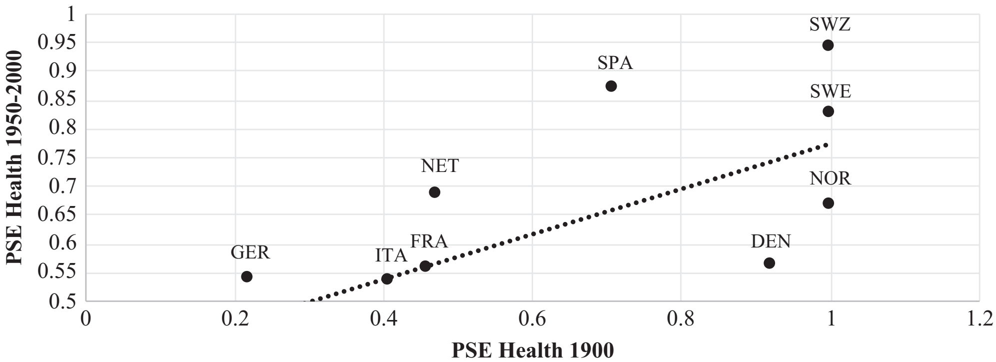
Source: Authors' calculations, see section 2.

Figure 6.1 PSE Education 1950−2010 and PSE Education 1900



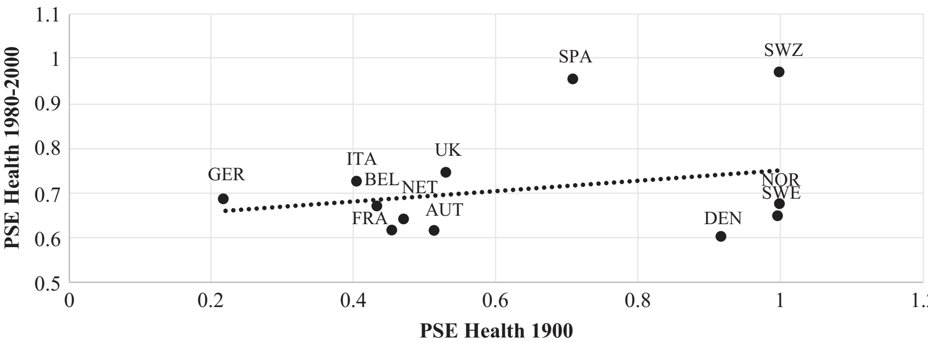
Source: Authors' calculations, see section 2.

Figure 6.2 PSE Education 1980−2010 and PSE Education 1900



Source: Authors' calculations, see section 2.

Figure 6.3 PSE Education 1950−2000 and PSE Education 1900



Source: Authors' calculations, see section 2.

Figure 6.4 PSE Education 1980–2000 and PSE Education 1900

3. A SIMPLE MODEL OF PUBLIC SECTOR EFFICIENCY AND INSTITUTIONS

We use a simple model to illustrate the link between public sector efficiency and core institutional fundamentals. In this model, the historical experience of institutional quality affects public sector efficiency, in an environment where institutional quality shapes public sector employee behaviour and in turn the efficiency of the public sector, whilst being determined by the behaviour it has influenced over time. To derive analytical results, we choose simplified functional forms and work in partial equilibrium, given factor prices and government revenue.

3.1 Public Sector Performance, Efficiency and Institutions

Assume that the public good is produced by bureaucrats, $i = 1, 2, \ldots, N$, public sector employees who work in the public sector, and who we assume are identical. The public sector produces the public good y_t, in period $t = 1, \ldots, T$, by using the effort of bureaucrats according to a linear production function:

$$y_t = A_t L_t. \tag{6.1}$$

The bureaucrats' contribution is given by $L_t = \Sigma_{i=1}^{N} l_t^i = Nl_t$, where l_t^i is the effort of bureaucrat i. The term A_t determines productivity and is affected by the country's experience of institutional quality, $I_0, .., I_t \in [0, 1]$, where I_0 represents an initial condition. In particular, we assume that:

$$A_t = a_t\left(I_0, .., I_t\right) \in [\underline{A}, \overline{A}],\ \underline{A} > 0 \tag{6.2}$$

where $a_t\,(I_0, .., I_t)$ is an increasing function and embodies the influence, over time, of institutional factors such as the protection of property rights, quality of rule of law, monitoring of government officials and public management. These factors define the structural framework within which the bureaucrats, as public sector employees, work, thus determining their productivity. Note that the function $a_t\,(I_0, .., I_t)$ can change over time, thus allowing the influence of historical experience to also change over time. For example, we can think of a_t as a weighted average of the elements in $(I_0, .., I_t)$, with the weights being time dependent to reflect differences in importance of past versus contemporary experience over time, or to allow for specific years, or events, to carry disproportionate weight in defining the institutional history. The persistence of the effect of institutional quality on A_t implies, in turn, persistence in A_t itself over time, and thus persistence in public sector performance.

The cost of producing the public good is given by w_tL_t, where $w_t > 0$ reflects the public sector wage policy, and is covered by the general government budget, which is given by:

$$w_t L_t + \left(1 - I_t\right) R_t = R_t + K_t, \tag{6.3}$$

where $R_t \geq 0$ denotes government revenue whose allocation is affected by institutional quality, and $K_t > 0$ is government revenue that is not subjected to potential misallocation. The left-hand side of (6.3) captures the elements of public expenditure. In this specification, the term $(1 - I_t)$ R_t defines the slack in government spending that results from poor institutional quality and implies that a share of public spending does not contribute to the production of the public good, but is instead diverted in uses that do not promote public sector output. In other words, poor institutional quality leads to a misallocation of national resources. We take R_t as given, which implies that we do not study the negative effects of institutional quality on aggregate output,[8] and instead we focus on its effects on the allocation of public funds.

The quality of institutions is an endogenous quantity at the aggregate level, depending on the actions of the government officials. In particular, we assume that I_t is determined by:

$$I_t = f_t\left(l_t, H_t\right), \tag{6.4}$$

where $f_t(l_t, H_t)$ is increasing in both inputs and:

$$H_t = h_t\left(I_0, .., I_{t-1}\right) \in [0,1], \tag{6.5}$$

is an increasing function capturing persistence in institutional quality. As with the productivity function $a_t\,(I_0, .., I_t)$, and for the same reasons, $h_t\,(I_0, .., I_{t-1})$ can change over time. For example, if $h_t\,(I_0, .., I_{t-1})$ is a weighted average of past institutional quality with the requirement that the weights sum to one, allowing the weights to change over time can capture differences in the importance of specific periods or can determine how quickly past influence fades. The specification in (6.4)–(6.5) then captures the idea that, conditional on the effect of past institutional quality, the more public sector employees work to perform their duties on average, the greater the improvement in institutional quality. The importance of current bureaucratic effort relative to past experience is also allowed to be time dependent: for example, if $f_t\,(l_t, H_t)$ is a weighted average of its two inputs, with the requirement that the weights sum to one, the weights can be time dependent. The general specifications in (6.2) and (6.4)–(6.5) allow the model to encompass a variety of possible trajectories, in terms of persistence of institutions and public sector performance, for different countries.

We define public sector efficiency to be a function of two inputs, public sector output, A_tL_t and public sector expenditure, $w_tL_t + (1 - I_t)R_t$:[9]

$$PSE_t = g\left(\left(A_t L_t\right), \left(w_t L_t + \left(1 - I_t\right) R_t\right)\right), \tag{6.6}$$

where g(·,·) is increasing in the first and decreasing in the second input. As an example, we can consider the ratio of public sector output to expenditure, that is:

$$PSE_t = \frac{A_t L_t}{w_t L_t + (1 - I_t) R_t}. \tag{6.7}$$

The importance of the quality of institutions is reflected in the term $(1 - I_t)R_t$, which captures the public fund misallocation effect, in the term A_t, which captures the importance of the history of institutions in shaping the structure of the public sector that determines its productivity and, finally, via the term L_t, which captures the effort of the bureaucrats and is, in turn, a function of institutional quality.[10] We study this last channel in more detail next. Note that, as we have seen in Section 2, (6.6)–(6.7) imply that public sector efficiency may persist over time, in this formulation via the terms A_t and I_t, which are functions of past institutional quality.

3.2 Bureaucrats

Bureaucrats receive income from their public sector positions, but can also extract a share of public expenditure for their own benefit, via rent-seeking activities. Each agent *i* maximises utility:

$$u\left(c^i, u^i; y\right),$$

where c^i and u^i are respectively *i*'s consumption and leisure hours and *y* denotes the quantity of public good. The utility function, in each period *t*, is given by:

$$u\left(c_t^i, u_t^i; y_t\right) = \ln c_t^i + \mu \ln u_t^i + \nu \ln y_t, \tag{6.8}$$

where μ, $v > 0$ are preference parameters. The time constraint of each *i* is:

$$l_t^i + s_t^i + u_t^i = 1, \tag{6.9}$$

where l_t^i and s_t^i are respectively *i*'s effort hours allocated to work and rent-seeking activities. The budget constraint of agent *i* is:

$$c_t^i = w_t l_t^i + \left(\frac{s_t^i}{S_t}\right)(1 - I_t)R_t, \tag{6.10}$$

where $w > 0$ determines the bureaucrat's wage as a public sector employee.

The term $(1 - I)R$ captures the part of government revenue diverted away from public good provision that defines a contestable prize for rent seekers.[11] Our modelling captures the idea that a lower quality of institutions allows bureaucrats to appropriate a share of national resources under the government's control. For example, the bureaucrats can divert national resources to activities that do not promote public good provision, but instead may relate to self-interest, and/or can lead to an increase in their own resources. We assume that the public sector employees can appropriate a share $\left(\frac{s_t^i}{S_t}\right)$ of the contestable prize. In particular, public sector employees compete with one another to extract a higher share of the pie, and this is captured by a typical Tullock (1967, 1980) type rent-seeking redistributive contest. In this specification,

$S_t \equiv \Sigma_i s_t^i$ is the sum of rent-seeking activities, so that $\left(\frac{s_t^i}{S_t}\right)$ is a standard rent-seeking technology that gives the fraction of the contestable prize extracted by each *i*.[12]

Each agent *i* chooses c_t^i, l_t^i, s_t^i, to maximise (6.8) subject to (6.9) and (6.10), taking the sum of rent-seeking activities, S_t, and institutional quality, I_t, as given. As shown in Appendix A, the choice of l_t is increasing in I_t.

3.3 Equilibrium

We focus on a partial equilibrium given sequences of functions $\{a_t, h_t, f_t\}_{t=1}^{T}$, time series for government revenue $\{R_t\}_{t=1}^{T}$ and payments $\{w_t\}_{t=1}^{T}$ and an initial level of institutional quality I_0. In a symmetric equilibrium, where all agents make the same choices, we drop the individual specific superscripts, so that, in each time period, $L_t \equiv \Sigma_i\, l_t^i = Nl_t$, $S_t \equiv \Sigma_i\, s_t^i = Ns_t$ and the optimality conditions for c_t, l_r and s_t, as determined by the bureaucrat's problem, are given by:

$$\frac{\mu}{(1-l_t-s_t)} = \frac{w_t}{c_t}, \tag{6.11}$$

$$\frac{\mu}{(1-l_t-s_t)} = \frac{(1-I_t)R_t}{c_t S_t}, \tag{6.12}$$

$$c_t = w_t l_t + \left(\frac{s_t}{S_t}\right)(1-I_t)R_t, \tag{6.13}$$

where:

$$I_t = f_t\left(l_t, H_t\right). \tag{6.14}$$

This makes clear that, in each period *t*, in equilibrium, a bureaucrat's effort l_t and the quality of institutions, I_t, are jointly determined, in conjunction with c_t and s_t. In turn, l_t and I_t determine K_t via (6.3), whereas l_t, I_t and $a_t\,(I_0, .., I_t)$ determine y_t and PSE_t, via (6.1) and (6.6) respectively. In Appendix A, we provide a formal definition of the equilibrium and show that under mild regularity conditions a unique equilibrium exists.

3.4 Predictions

In equilibrium, (6.6) shows channels for public sector efficiency PSE_t to increase with prior good history of quality of institutions, which, in period *t* is pre-determined. Good institutional quality in the past can affect public sector efficiency in the present via two channels: public sector performance and the expenditure required to achieve public sector outputs. For both channels, the effort of the public sector employees is a key ingredient. As can be seen in (6.6), an improved past institutional quality increases the productivity of public sector employees, via the term A_t, increasing public sector output and public sector performance indicators, which should also increase public sector efficiency. Moreover, via the term $(1 - I_t)\, R_t$, it implies a smaller misallocation of public sector funds to less productive uses, which reduces wasteful public sector expenditure and also, other things equal, tends to increase public sector efficiency. Finally, it increases public sector employee effort, l_t, which is increasing in I_t, in turn increasing both public sector output and expenditure. If the improvement of public sector output outweighs the effect of the additional expenditure, then a better past institutional quality tends, other things equal, to increase public sector efficiency via this term as well.

The model analysis implies that exogenous changes that led to better institutional quality in the past should have a bearing on current public sector performance and efficiency, via different channels, which may have different strengths empirically. However, although these channels can lead to a positive effect from past institutional quality on future public sector performance and efficiency outcomes, the overall relationship between public sector efficiency and the quality of past institutions empirically also depends on a range of policies and other factors that may affect the time series of $\{R_t\}_{t=1}^{T}$ and $\{w_t\}_{t=1}^{T}$, and the functional forms of $\{a_t, h_t, f_t\}_{t=1}^{T}$. For example, political events or technological change that may reduce the dependence of contemporary institutions on past experience (that is, changes in $\{a_t, h_t, f_t\}_{t=1}^{T}$ over time such that a lower weight is attached to past values of institutions), or policies that affect public sector wages and expenditure (that is, changes in $\{R_t\}_{t=1}^{T}$ and $\{w_t\}_{t=1}^{T}$), can amplify or dampen the effect of past institutions on contemporary public sector efficiency.

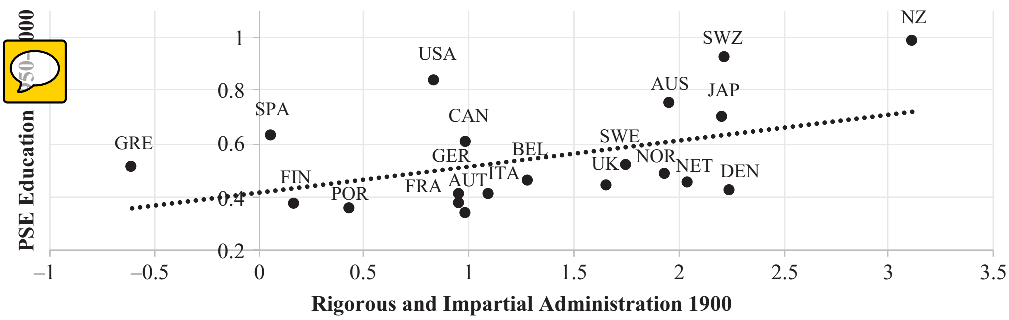
4. THE ROLE OF EARLY INSTITUTIONAL QUALITY

The simple model we developed in the previous section places the spotlight on the role of institutions as a factor that can have long-lasting effects on Public Sector Efficiency (PSE), and thus lead to persistency in PSE. In this section, we explore the relationship between past institutional quality and future PSE. More precisely, we focus on institutions that can affect public sector performance via increased public sector output, and/or misallocation of public funds and of the effort of government officials via rent seeking. We investigate the potential effects of: (i) a rigorous and impartial public administration, also described as Weberian state bureaucracy by the relevant literature (see for example Coppedge *et al.* (2019); Cornell *et al.* (2020)) and (ii) the existence of well-defined property and contract rights and a well-established Rule of Law which ensures that the laws are transparently, independently and equally enforced, and that the government officials comply with the legal framework (see Coppedge *et al.* (2019), for more details on this).

The Weberian ideal type of bureaucracy entails hierarchical organisation with clearly delineated lines of authority and areas of responsibility, and decisions that are based on clearly codified rules and made in an impartial manner (Weber, 1978).[13] A solid and impartial public administration, a Weberian state bureaucracy, is a characteristic that describes how the state administration is organised and is distinguished from other outcome-oriented concepts such as ‘state capacity’ or ‘quality of government’.[14] According to the relevant literature (see for example Evans and Rauch, 1999; Cornell *et al.* 2020) a solid and impartial public administration is expected to increase competence among bureaucrats − which in turn reduces the bureaucratic slack − and ensures enforcement of property and contract rights. Therefore, the existence of a solid and rigorous public administration is expected to affect PSE both through increasing public sector productivity, and by decreasing rent seeking activities − which imply misallocation of public funds and of government officials’ efforts from producing the public sector output. Similarly, the existence of well-established Rule of Law which ensures that the laws are transparently, independently and equally enforced among citizens and that the government officials comply with the legal framework is expected to affect PSE via increasing public sector productivity and decreasing rent-seeking activities.

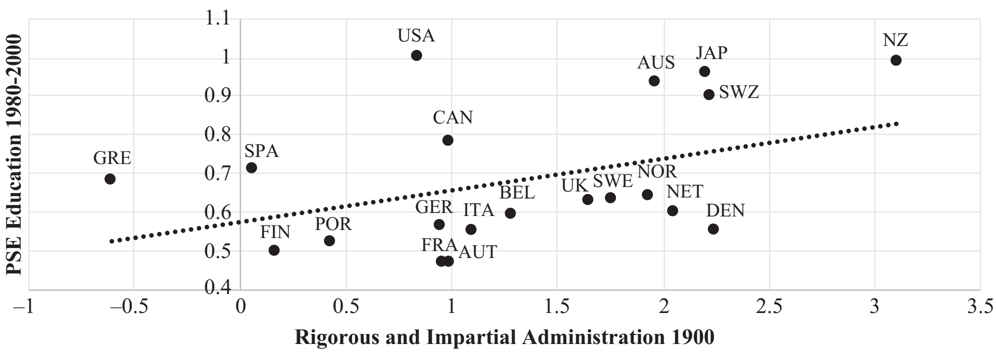
The scatterplots in Figures 6.5–6.8 show a positive association between alternative proxies of the Quality of Institutions at the turn of the twentieth century (that is Rule of Law or Rigorous and Impartial Public Administration) and PSE in Education during the 1950–2010 period or the 1980–2010 period. More precisely, Figures 6.5–6.6 present a positive relationship between PSE scores and the Quality of Institutions in 1900 when the latter is proxied by Rigorous and Impartial Public Administration and Figures 6.7–6.8 show again a positive − although weaker − relationship between PSE scores and the Quality of Institutions in 1900 when the latter is proxied by Rule of Law. The difference that we observe between these two alternative measures of institutional quality can potentially be attributed to the fact that at the turn of the twentieth century most of the countries in our sample had already achieved relatively satisfactory levels of Rule of Law (and therefore there is no sufficient cross-country variation) whereas in the case of Rigorous and Impartial Public Administration the cross-country variation is larger and this reveals better the positive association.

A similar message is derived from Figures 6.9–6.12 that present scatterplots where in the horizontal axis we measure the Quality of Institutions during the early twentieth century (that is Rule of Law or Rigorous and Impartial Public Administration) and in the vertical axis the



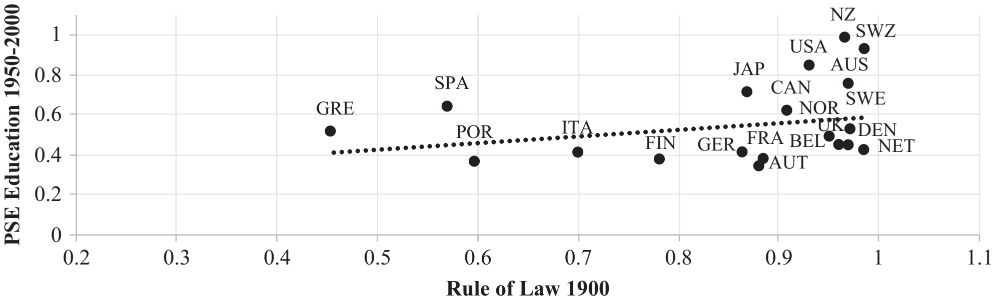
Source: Authors' calculations, see section 2; Coppedge *et al.* (2019).

Figure 6.5 PSE Education 1950−2010 and Quality of Institutions 1900



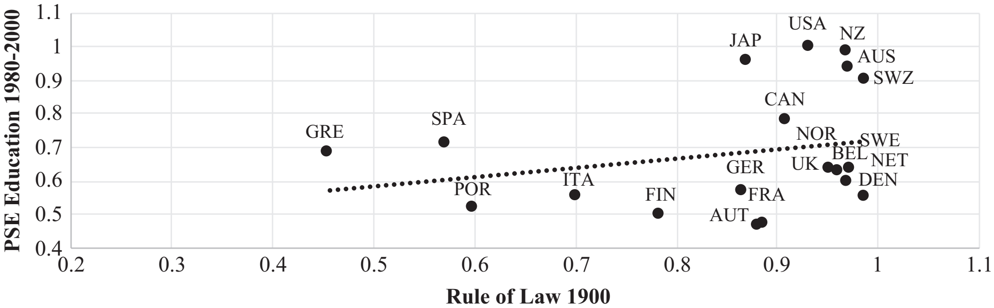
Source: Authors' calculations, see section 2; Coppedge *et al.* (2019).

Figure 6.6 PSE Education 1980−2010 and PSE Education 1900



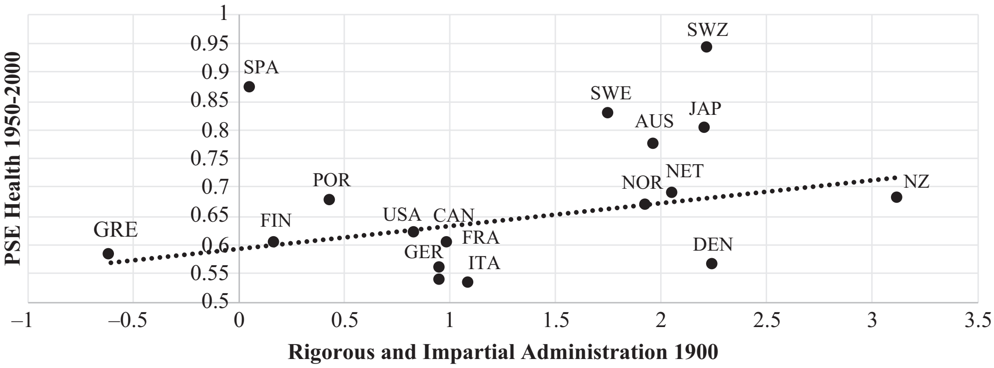
Source: Authors' calculations, see section 2; Coppedge *et al.* (2019)

Figure 6.7 PSE Education 1950–2010 and PSE Education 1900



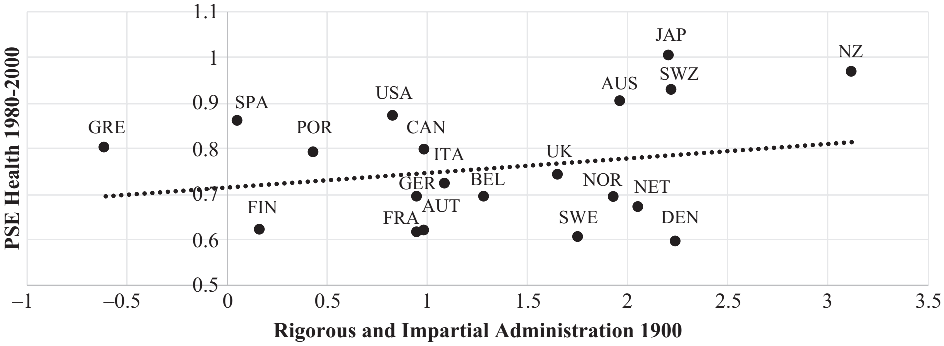
Source: Authors' calculations, see section 2; Coppedge *et al.* (2019).

Figure 6.8 PSE Education 1980–2010 and PSE Education 1900



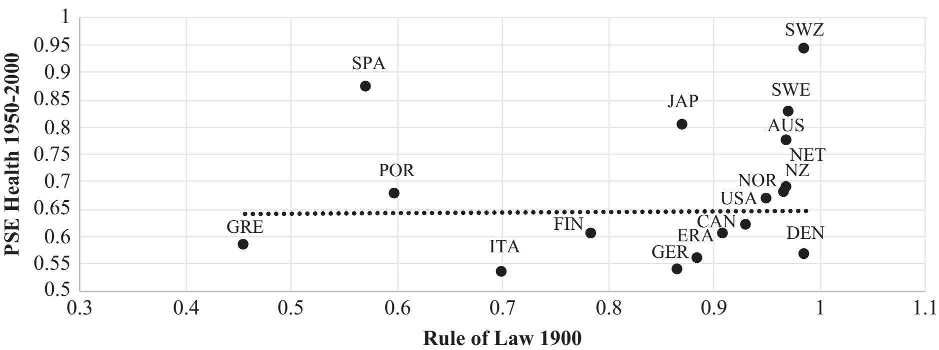
Source: Authors' calculations, see section 2; Coppedge *et al.* (2019).

Figure 6.9 PSE Education 1950–2010 and Quality of Institutions 1900



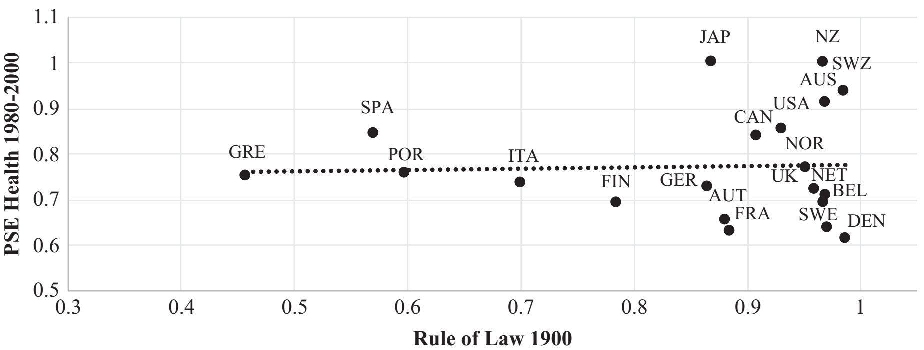
Source: Authors' calculations, see section 2; Coppedge *et al.* (2019).

Figure 6.10 PSE Education 1980–2010 and Quality of Institutions 1900



Source: Authors' calculations, see section 2; Coppedge *et al.* (2019).

Figure 6.11 PSE Education 1950–2010 and Quality of Institutions 1900

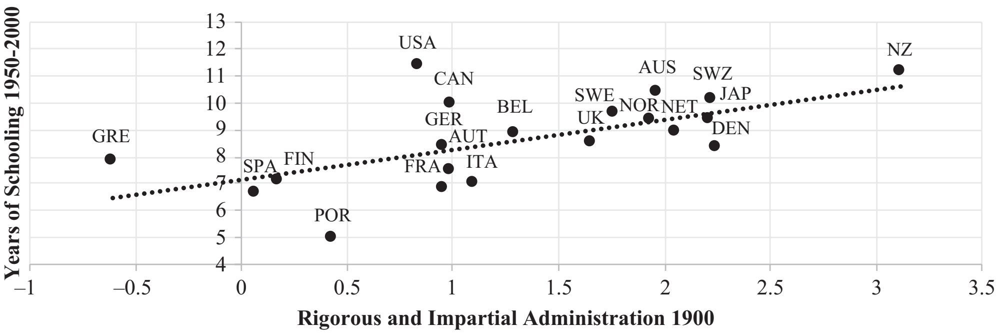


Source: Authors' calculations, see section 2; Coppedge *et al.* (2019).

Figure 6.12 PSE Education 1980–2010 and Quality of Institutions 1900

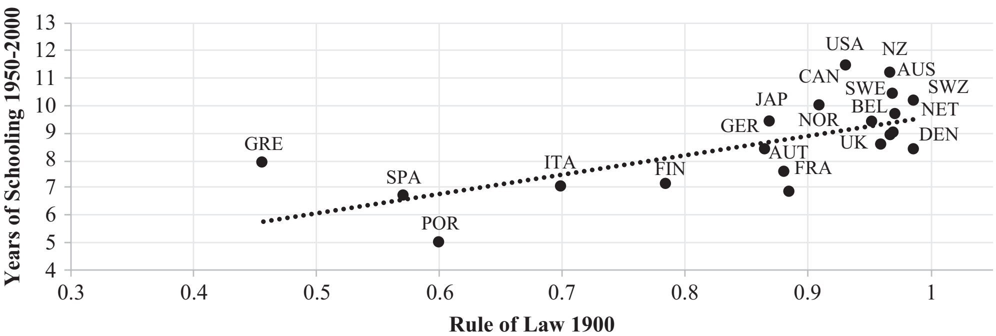
PSE in Health. As before, Figures 6.9–6.10 present a positive association between PSE scores and the Quality of Institutions in 1900 when the latter is proxied by Rigorous and Impartial Public Administration (and so there is sufficient cross country variation in our sample). In contrast, in Figures 6.11–6.12 where the Quality of Institutions is captured by Rule of Law the scatterplots show a weak association between the variables under examination.

We further investigate the relationship between the historical experience of institutional quality in a country and future public sector performance. To this end, we present scatterplots where in the horizontal axis we measure alternative proxies of historical institutional quality (that is Rule of Law in 1900 or Rigorous and Impartial Public Administration in 1900) and in the vertical axis we measure now PSP in Education (that is average years of schooling) and PSP in Health (that is infant mortality) during the 1950–2010 period. As can be seen, Figures 6.13–6.16 suggest a positive association between historical institutional quality and alternative Public Sector Performance (PSP) measures, which is clearer than the relationship in Figures 6.9–6.12.[15] This positive association indicates that, quantitatively, improvements via public sector performance is a strong channel for institutional quality to affect public sector efficiency.



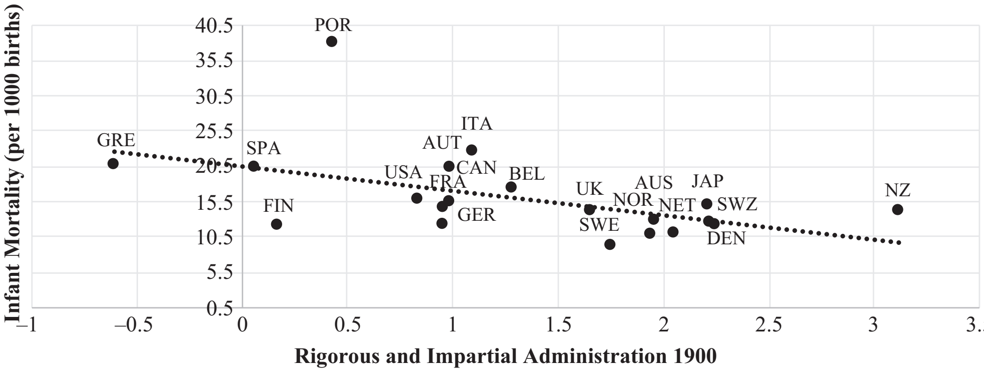
Source: Authors' calculations, see section 2; Coppedge *et al.* (2019).

Figure 6.13 Years of Schooling 1950−2010 and Quality of Institutions 1900



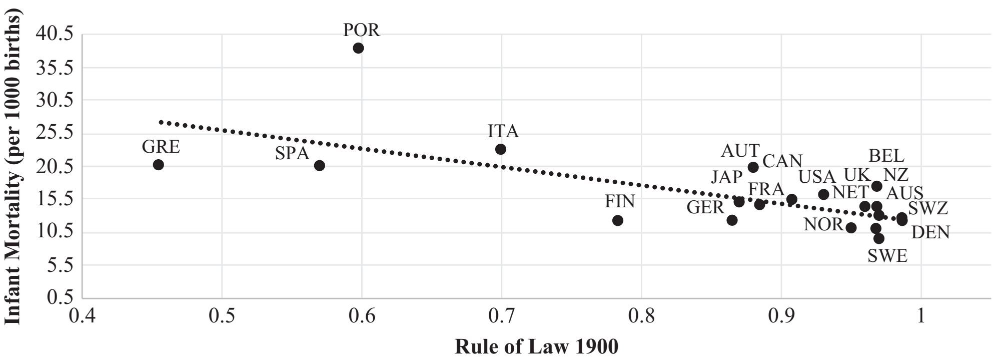
Source: Authors' calculations, see section 2; Coppedge *et al.* (2019).

Figure 6.14 Years of Schooling 1950−2010 and Quality of Institutions 1900



Source: Authors' calculations, see section 2; Coppedge *et al.* (2019).

Figure 6.15 Years of Schooling 1950−2010 and Quality of Institutions 1900



Source: Authors' calculations, see section 2; Coppedge *et al.* (2019).

Figure 6.16 Years of Schooling 1950−2000 and Quality of Institutions 1900

5. CONCLUSIONS

In this chapter, we studied the historical origins of public sector efficiency, and explored its links with institutional quality. We focused on public sector efficiency as approximated by performance in education and health, which have been within the set of policy objectives for governments throughout the twentieth century, and can be consistently measured for OECD countries over this period. We focused on institutions that can affect public sector performance via public sector output, and misallocation of public funds and of the effort of government officials via rent seeking.

We found that comparative public sector efficiency persists, and that institutional quality in the distant past can affect public sector performance and efficiency persistently, even a century later. In particular, we find that the effect of historical experience of institutions is stronger on public sector performance, compared with public sector efficiency.

The implication of these results is that exogenous changes that lead to better institutional quality today should have a bearing on future public sector performance and efficiency in the medium run to long run. Examples of such positive changes include public sector policies that protect national resources from rent seeking and political activity that can lead to misallocation of public funds to areas that are not directly related to public good provision. However, we have not investigated whether there is a short-run effect, and further research is needed to investigate a possible causal contemporaneous effect. This should be of immediate importance for policy interventions aiming to have a more rapid effect.

A caveat of our analysis is that we focused on health and education as targets of public intervention. This was motivated by data availability, but also because of the need to examine areas that have been over time associated with public sector interventions. In more recent decades, the public sector has assumed more roles and objectives, and thus the factors that can effect a change in public sector efficiency, as well as the channels via which this may happen, have also increased. Therefore, to understand how institutional reform may improve public sector efficiency today, the additional complexities of the current public sectors need to be accounted for. We believe, however, that our study has demonstrated the potential that improvements in institutional quality have to generate positive change.

APPENDIX A

A symmetric equilibrium is an allocation $\{c_t, s_t, l_t, K_t, y_t, I_t, PSE_t\}_{t=1}^{T}$ that solves (6.1), (6.14), (6.3), (6.7), (6.11), (6.12) and (6.13), where $L_t = Nl_t$ and $S_t = Ns_t$, given an initial condition I_0, sequences of functions $\{a_t, h_t, f_t\}_{t=1}^{T}$, and time series for government revenue $\{R_t\}_{t=1}^{T}$ and payments $\{w_t\}_{t=1}^{T}$ that satisfy the condition $R_t < \frac{Nw_t}{(1+\mu)}$, for each t.[16] In Proposition 1 we establish existence and uniqueness of equilibrium.

Proposition 1

If $f_t(l_t,\cdot)$ is continuous and increasing, an equilibrium exists. If $f_t(l_t,\cdot)$ is also either concave over its domain, or convex over its domain, the equilibrium is unique.

Proof

For each time period t, the first order conditions from the problem of the bureaucrat imply that the time allocated to rent seeking is:

$$s_t = \frac{(1-I_t)R_t}{Nw_t}, \tag{6.15}$$

which is decreasing in I_t. Note that $s_t \in [0, \frac{R_t}{Nw_t}] \subset [0, \frac{1}{(1+\mu)}) \subset [0,1]$, for $R_t < \frac{Nw_t}{(1+\mu)}$. In turn, the effort allocated to productive work is:

$$l_t = \frac{1}{(1+\mu)} - \frac{(1-I_t)R_t}{Nw_t}, \tag{6.16}$$

which is increasing in I_t and $l_t \in \left[\frac{1}{(1+\mu)} - \frac{R_t}{Nw_t}, \frac{1}{(1+\mu)}\right] \subset (0, \frac{1}{(1+\mu)}] \subset [0,1]$.

An equilibrium in a given year t exists if a solution to (6.16) can be found. The equilibrium allocations $\{c_t, s_t, l_t, K_t, y_t, I_t, PSE_t\}_{t=1}^{T}$ can then be found recursively from period t = 1 forward. The *lhs* of (6.16) defines the 45^0 line over $\left[\frac{1}{(1+\mu)} - \frac{R_t}{Nw_t}, \frac{1}{(1+\mu)}\right]$. A solution to (6.16) is then a crossing point of the function:

$$g(l) \equiv \frac{1}{(1+\mu)} - \frac{(1-f_t(l_t, H_t))R_t}{Nw_t},$$

and the 45^0 line in $\left[\frac{1}{(1+\mu)} - \frac{R_t}{Nw_t}, \frac{1}{(1+\mu)}\right]$. The function $g(l)$ is continuous and increasing in l. Because $f_t(l_t,H_t)$ is (strictly) increasing, and $f_t(0, H_t) \geq 0$, $f_t\left(\frac{1}{(1+\mu)} - \frac{R_t}{Nw_t}, H_t\right) > 0$, which implies that $g\left(\frac{1}{(1+\mu)} - \frac{R_t}{Nw_t}\right) > \frac{1}{(1+\mu)} - \frac{R_t}{Nw_t}$. Because $f_t\left(\frac{1}{(1+\mu)}, H_t\right) \geq 0$, $g\left(\frac{1}{(1+\mu)}\right) < \frac{1}{(1+\mu)}$. Therefore, for each t, $g(l)$ crosses the 45^0 line in $\left[\frac{1}{(1+\mu)} - \frac{R_t}{Nw_t}, \frac{1}{(1+\mu)}\right]$, implying that an equilibrium exists. When $g(l)$ is either only concave or only convex over $\left[\frac{1}{(1+\mu)} - \frac{R_t}{Nw_t}, \frac{1}{(1+\mu)}\right]$, there is a unique crossing point so that there is a unique solution to (16), implying that the equilibrium is unique. ■

NOTES

1. See for example North (1990), Persson and Tabellini (2000), Drazen (2000), Besley and Persson (2009), Besley and Ghatak (2010) and Acemoglu and Robinson (2012, 2019).
2. We measure public spending net of debt repayment expenditure, because during the historical periods that we study the latter have sometimes been a significant burden on the government budget, reflecting exogenous episodes, without relating to objectives relevant to public good provision.
3. Compared with other measures of public sector performance in the literature, we do not use measures of infrastructure. Such measures of economic activity, that are appropriate to relate to the public sector since the turn of the twentieth century, are hard to find.
4. Primary government expenditure (as a per cent of GDP) equals total government expenditure (per cent of GDP) minus interest paid on public debt (per cent of GDP).
5. Alternative data of government spending (for example Flora *et al.* (1983), Mitchell (2013), Lindert (1994, 2004)) do not follow a common classification among countries.
6. DEA is a deterministic, linear programming technique that provides a piecewise frontier, by enveloping the observed data points, and yields a convex production possibilities set.
7. Using decade averages, instead of higher frequency data, is preferable for two main reasons. First, the effect of various policies need not be immediate, but instead may take time to generate results. Second, we want to exclude potential business cycle effects from the input of our analysis (that is primary government spending).
8. More generally, R can be thought of as a function of the economy's income that the government can tax. Here we take the size of the government budget as given and focus on its allocation.
9. Here, we are interested in public sector expenditure efficiency, as opposed to public sector efficiency in terms of productivity, which would be defined in terms of labour input.
10. Note that we could add I_t in the numerator of (6.7), since this is another output of public sector bureaucrats. However, since in our empirical analysis we investigate the relationship between institutional quality and public good provision, which is captured here by y_t, we omit I_t from the numerator of (6.7).
11. See for example Angelopoulos *et al.* (2021) for references to the literature on formulations of the contestable prize for rent-seeking activity as a function of national resources.
12. See for example Murphy *et al*. (1991), Dixit (2004, chapter 5), Hillman (2009, chapter 2), Angelopoulos *et al.* (2009, 2011, 2021) and Esteban and Ray (2011) for similar rent-seeking technologies.
13. Moreover, in a Weberian ideal type of bureaucracy bureaucrats are meritocratically recruited, have expert training, and advance in the organisation based on objective criteria.
14. This characteristic is important for the purposes of our study. This is because these outcome-centred concepts (for example corruption or quality of government) have been widely employed by previous PSE studies as PSP indicators (see for example Afonso *et al.* (2005); Adam *et al.* (2011)).
15. The negative association between historical institutional quality and infant mortality presented in Figures 6.15 and 6.16 implies positive association with the inverse of infant mortality, which is the relevant Public Sector Performance measure that we employ (see Table 6.1 for more details on this).
16. The condition $R_t < \frac{Nw_t}{(1+\mu)}$ requires that the contestable prize (misallocation of public funds) is sufficiently small to permit well defined time allocations (see the proof of Proposition 1).

REFERENCES

Acemoglu, D. and J. Robinson (2012). *Why nations fail: The origins of power, prosperity, and poverty*, Profile Books.
Acemoglu, D. and J. Robinson (2019). Rents and economic development: The perspective of Why Nations Fail, *Public Choice*, 181, 13–28.
Adam, A., M. Delis and P. Kammas (2011). Public sector efficiency: Leveling the playing field between OECD countries, *Public Choice*, 146 (1), 163–183.
Afonso A., J. Jalles and A. Venancio (2020). Government Spending Efficiency, Measurement and Applications: A Cross-country Efficiency Dataset, *Working Papers* REM 2020/0147, ISEG - Lisbon School of Economics and Management, REM, Universidade de Lisboa.
Afonso, A. and M. Kazemi (2016). Assessing Public Spending Efficiency in 20 OECD Countries, *Working Papers* Department of Economics 2016/12, ISEG - Lisbon School of Economics and Management, Department of Economics, Universidade de Lisboa.
Afonso, A., L. Schuknecht and V. Tanzi (2005). Public sector efficiency: An international comparison, *Public Choice*, 123, 321–347.
Afonso, A., L. Schuknecht and V. Tanzi (2010). Public sector efficiency: Evidence for new EU member states and emerging markets, *Applied Economics*, 42, 2147–2164.

Afonso, A. and M. St. Aubyn (2005). Non-parametric approaches to education and health efficiency in OECD countries, *Journal of Applied Economics*, 8, 227–246.
Afonso, A. and M. St. Aubyn (2006). Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs, *Economic Modelling*, 23, 476–491.
Angelopoulos, K., A. Philippopoulos and V. Vassilatos (2009). The social cost of rent seeking in Europe, *European Journal of Political Economy*, 25, 280–299.
Angelopoulos, K., G. Economides and V. Vassilatos (2011). Do institutions matter for economic fluctuations? Weak property rights in a business cycle model for Mexico, *Review of Economic Dynamics*, 14, 511–531.
Angelopoulos, A., K. Angelopoulos, S. Lazarakis and A. Philippopoulos. (2021). The distributional consequences of rent seeking, *Economic Inquiry*, 59(4), pp. 1616–1640.
Besley, T. and T. Persson (2009). The origins of state capacity: Property rights, taxation and politics, *American Economic Review*, 99, 1218–1244.
Besley, T. and M. Ghatak (2010). Property rights and economic development, in D. Rodrik and M. Rosenzweig (editors), *Handbook of Development Economics*, vol. 5, Elsevier.
Coppedge, M., J. Gerring, C. H. Knutsen, S. Lindberg, J. Teorell, D. Altman, M. Bernhard, S. Fish, A. Glynn, A. Hicken, A. Ltihrmann, K. Marquardt, K. McMann, P. Paxton, D. Pemstein, B. Seim, R. Sigman, S. E. Skaaning, J. Staton, S. Wilson, A. Cornell, L. Gastaldi, H. Gjerl0w, N. Ilchenko, J. Krusell, L. Maxwell, V. Mechkova, J. Medzihorsky, J. Pernes, J. von Romer, N. Stepanova, A. Sundstrom, E. Tzelgov, Y. Wang, T. Wig, and D. Ziblatt (2019). V-Dem [Country-Year/Country-Date] Dataset v9.0, Varieties of Democracy (V-Dem) Project.
Cornell, A., C. H. Knutsen and J. Teorell (2020). Bureaucracy and growth, *Comparative Political Studies,* 53(14), 2246–2282.
Dixit, A. (2004). *Lawlessness and Economics*, Princeton University Press.
Drazen, A. (2000). *Political Economy in Macroeconomics*, Princeton University Press.
Esteban, J. and D. Ray (2011). Linking conflict to inequality and polarization, *American Economic Review*, 101, 1345–1374.
Evans, P. and J. E. Rauch (1999). Bureaucracy and growth: A cross-national analysis of the effects of 'Weberian' state structures on economic growth, *American Sociological Review*, 64, 748–765.
Flora, P., J. Alber, R. Eichenberg, J.K.F. Krausm, W. Pfenning, and K. Seebohm (1983). *State, Economy and Society 1815–1975*, Campus Verlag, Frankfurt Germany.
Hillman, A. (2009). *Public Finance and Public Policy: Responsibilities and Limitations of Government* (2nd ed.), Cambridge University Press.
Lee, J. W. and H. Lee (2016). Human capital in the long run. *Journal of Development Economics*, 122, 147–169.
Lindert, P. (1994). The rise in social spending 1880–1930. *Explorations in Economic History*, 31, 1–37.
Lindert, P. (2004). *Growing Public. Social Spending and Economic Growth Since the Eighteenth Century*, Cambridge University Press.
Mauro, P., R. Romeu, A. Binder and A. Zaman (2015). A modern history of fiscal prudence and profligacy, *Journal of Monetary Economics*, 76, 55–70.
Mitchell, B.R. (2013). *International Historical Statistics: Europe, 1750–2013*, Palgrave Macmillan, London, UK
Murphy, K., A. Shleifer and R. Vishny (1991). The allocation of talent: Implications for growth, *Quarterly Journal of Economics*, 106, 503–530.
North, D. (1990). *Institutions, Institutional Change and Economic Performance*, Cambridge University Press.
Persson, T. and G. Tabellini (2000). *Political Economics*, MIT Press.
Tullock, G. (1967). The welfare costs of monopolies, tariffs and theft, *Western Economic Journal*, 5, 224–232.
Tullock, G. (1980). Efficient rent-seeking, in J. M. Buchanan, R. D. Tollison and G. Tullock (editors), *Toward a Theory of the Rent-Seeking Society*, Texas A&M University Press, College Station.
Weber, M. (1978). *Economy and society: An outline of interpretive sociology*, University of California Press.
World Health Organization (WHO) (2020). World Health Data Platform https://www.who.int/data

7. Government efficiency and fiscal rules

Amélie Barbier-Gauchard, Kéa Baret, and Xavier Debrun

1. INTRODUCTION

Fiscal rules have been on the rise worldwide since the 1990s, highlighting the urgent need for sound public finance. Since the turn of the century, furthermore, the issue of government efficiency has stimulated public debate and given rise to numerous indicators. Government efficiency refers to their ability to reach economic and social objectives with public funds but with the lowest possible costs. The fields of fiscal discipline and government efficiency have been developing in parallel but independently, although the two fields seem related and probably should be considered together. Indeed, does government efficiency lead to better fiscal discipline? Or, on the contrary, do fiscal rules make governments more efficient?

'Fiscal discipline' refers to a government's ability to maintain sound public finance. A first strand of literature considers that the effectiveness of fiscal rules should be measured by their disciplinary effect on fiscal behavior since the seminal paper of Bohn and Inman (1996). A second strand of literature considers that the performance of fiscal rules should be measured by their impact on various macroeconomic aggregates (activity, employment, inflation, interest rates, terms of trade and so on) as underlined, for instance, by Ghosh et al. (2011). A third strand of literature has focused on the key determinants of fiscal rule compliance (Reuter, 2019; Larch and Santacroce, 2020; Barbier-Gauchard et al., 2021b).

Meanwhile, as underlined by Afonso et al. (2021b), 'government efficiency' encompasses the public sector's ability to reach objectives funded by public money at the lowest possible cost to taxpayers. The most in-depth measure of government efficiency is the 'Public Sector Performance (PSP)' indicator proposed by Afonso et al. (2005) in their seminal article. This indicator has two main components: opportunity indicators and traditional Musgravian indicators. The opportunity indicators capture the institutional and economic context of the considered economy (corruption, shadow economy and so on) and the performance of several key public goods (education, health and infrastructure). The Musgravian indicators refer to the three functions of the public sector and assess income/wealth inequality (via the Gini coefficient), the stabilization power of the budget (deviations in GDP growth and inflation) and overall economic performance (GDP per capita, GDP growth and unemployment rate).

This large literature on fiscal rules does not contain any empirical evidence of a potential link between fiscal rules and government efficiency. In this context, this chapter investigates potential bridges between fiscal rules and government efficiency. More precisely, we investigate the correlation between the presence, strength or compliance rate of fiscal rules and a government efficiency index. This exploratory study offers some interesting insights but the conclusions depend on the indicator used as a proxy for fiscal rule performance, the periods considered and the types of fiscal rules included in the analysis.

The rest of this chapter is organized as follows. Section 2 discusses the potential relation between government efficiency and fiscal discipline. Section 3 proposes an overview of fiscal rules in the world. Section 4 offers an insight on the debate on fiscal rules performance. In section 5, the degree of correlation between different fiscal rule related proxies and government efficiency is tested. Section 6 concludes.

2. GOVERNMENT EFFICIENCY AND FISCAL DISCIPLINE: ORTHOGONAL OR JOINED AT BIRTH?

'Government efficiency' refers to the achievement of economic and social objectives typically funded by public funds with the lowest possible amount of taxpayer money. 'Fiscal discipline' can be defined as the ability to maintain sound public finance. What possible bridge can there be between these two objectives?

2.1 From Government Efficiency to Fiscal Discipline

On the one hand, as underlined by Afonso et al. (2021b), the concept of government efficiency encompasses the public sector's ability to reach economic and social objectives usually funded by public money at the lowest possible cost to taxpayers. A related concept is the necessary production (of public goods and services) at the lowest possible cost and as fast as possible. Government efficiency usually involves balanced budgets and ensuring appropriate and rational public spending with no waste. Since Musgrave's (1959) seminal paper, three functions have traditionally been entrusted to the government: i) allocation of public goods and services (education, health, security, defense and so on), ii) the redistribution of income from the richest to the poorest to improve social justice and iii) the cyclical stabilization of economic activity (in particular inflation and employment) following economic shocks.

It is worth noting that 'government efficiency' differs from 'government effectiveness', the public sector's ability to achieve its goals. 'Government effectiveness captures perceptions of the quality of public services, the quality of the civil service and its degree of independence from political pressure, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies', as explained by the World Bank and included among six measures of governance in its Worldwide Governance Indicators.[1] The government effectiveness index, calculated for 193 countries, runs from −2.5 for the least effective government to 2.5 for the most. This index has also generated an abundant literature focusing in particular on the determinants of government effectiveness and its relationship to the population's wellbeing (see García-Sánchez et al., 2013) or the association between this indicator and long-term growth (see for instance Alam et al., 2017).

The concept of 'government efficiency' also includes the ability to optimize available financial resources to achieve objectives. The most in-depth measure of government efficiency is the public sector performance (PSP) indicator proposed by Afonso et al. (2005) in their seminal article and explored in more detail in the present handbook. This indicator has two main components: opportunity indicators and traditional Musgravian indicators. The opportunity indicators capture the institutional and economic context of the considered economy (corruption, shadow economy and so on) and the performance of several key public goods (education, health and infrastructure). The Musgravian indicators refer to the three functions of the public sector mentioned above and assess income/wealth inequality (via the Gini coefficient), the stabilizing power of the budget (deviations in GDP growth and inflation) and overall economic performance (GDP per capita, GDP growth and unemployment rate). This index runs from 0.01356 (least effective) to 2.0557 (most effective). Table 7.1 summarizes the variables used to build the PSP indicator.

On the other hand, 'fiscal discipline' refers to a government's ability to maintain sound public finance. Fiscal discipline thus involves reducing deficit bias, increasing confidence in fiscal policy and reducing the costs of public borrowing to keep public debt sustainable. Fiscal discipline can be implemented through a wide range of measures, such as fiscal rules with numerical targets (balanced budget, public debt, expenditure, revenue and/or golden rules), independent fiscal institutions (generally called 'fiscal councils') in charge of monitoring public finance and medium-term budgetary frameworks. Numerical fiscal rules and medium-term frameworks differ in their implementation and time horizons. Medium-term budgetary

Table 7.1 Overall public sector performance indicators from a 'government efficiency' perspective

Sub Index	**Variable**
Opportunity Indicators	
Administration	Corruption
	Red Tape
	Judicial Independence
	Property Rights
	Shadow Economy
Education	Secondary School Enrolment
	Quality of Educational System
	Program for International Student Assessment (PISA) scores
Health	Infant Survival Rate
	Life Expectancy
	Cardiovascular Disease (CVD), Cancer, Diabetes or Chronic Respiratory Disease (CRD) Survival Rate
Public Infrastructure	Infrastructure Quality
Standard Musgravian Indicators	
Distribution	Gini Index
Stabilization	Coefficient of Variation of Growth
	Standard Deviation of Inflation
Economic Performance	GDP per Capita
	GDP Growth
	Unemployment

Source: Afonso et al. (2005).

frameworks allow governments to go beyond the annual budget cycle. Numerical fiscal rules are stricter and constrain fiscal discipline over shorter timeframes. Together, these tools constitute what is traditionally called the 'fiscal framework'.

2.2 Pros and Cons of Fiscal Discipline

As common-pool resources, public deficit and public debt biases require careful monitoring to avoid making public finance unsustainable (Wyplosz, 2013). Moreover, the problem of temporal inconsistency (Kydland and Prescott, 1977) creates a temptation for discretionary policies. This well-known bias fuels the famous 'rules versus discretion' debate, that is, whether or not public decision makers should be free to conduct their policies unchecked. Fiscal rules appear to be a useful solution to the problem of temporal inconsistency, in the sense that they also limit crowding-out effects. In a monetary union, a further strong argument for these rules is the presence of externalities linked to the close relations between countries: poor public finance management in any member state risks destabilizing the union as a whole, with a possible domino effect if a state goes bankrupt (see De Grauwe, 2000 or IMF, 2009).

However, if a fiscal rule is too strict, it can lead to a major cut in productive public spending and may therefore compromise growth. This point is particularly important in a monetary union without a substantial centralized budget as is the case in the European Monetary Union (EMU). For example, Thirlwall (2001), Warin (2005) and Wyplosz (2006) have argued that fiscal rules may have deteriorated economic growth in the EMU. These effects have been analyzed empirically at the national (Bohn and Inman, 1996) and supranational (EU Stability and Growth Pact; Castro, 2011) levels. Along the same lines, Gali and Perotti (2003) and Artis and Onorante (2006) have assessed the effects of fiscal rules on fiscal policy behavior, public investment and growth, finding that improvements in public fiscal balances are mainly due to higher economic growth and not to better government behavior.

Another crucial debate surrounds the pro-cyclicality of fiscal rules. While Ghosh et al. (2011) and Perotti (2013) found that fiscal rules do not prevent macroeconomic instability, Bova et al. (2014), Guerguil et al. (2017) (for developing countries) and Sacchi and Salotti (2015) (for OECD countries) note that the stabilizing power of fiscal rules depends on the type of rule implemented. The heterogeneous effects of different types of rules on fiscal discipline and economic stabilization raise the question of what the optimal choice is. As developed below, many variants of each type of rule are implemented in OECD countries.

In addition, Wyplosz (2013) and Debrun and Kumar (2007) discuss fiscal rules as driven by reverse causality: 'disciplined governments may wish to adopt rules as a way of cementing and signaling their determination'. This view is supported by Heinemann et al.'s (2018) metaregression analysis of data previously modeled by instrumental variable (Foremny, 2014), least squares dummy variable (Reuter, 2015), system-generalized method of moments (Bergman et al. (2016)) or propensity-score matching regression (Tapsoba, 2012; Barbier et al., 2021a). Finally, Wyplosz (2013) argues that voter preferences and other unobserved factors may affect the results of these studies, which nevertheless involved a large number of robustness tests to control for bias.

2.3 Potential Relationship between Fiscal Rules and Government Efficiency

These tools are designed to implement fiscal discipline. But what does a truly efficient fiscal discipline procedure consist of? And what is the true link between fiscal discipline and government efficiency? In particular, does fiscal discipline contribute to government efficiency? Or on the contrary, is it government efficiency that enables fiscal discipline? The two are intertwined it seems; however, the literature on these topics is tightly compartmentalized.

On the one hand, the fiscal discipline literature tries to define what efficient discipline is and what the best public finance indicators/rules to monitor it could be. Clearly, a fiscal discipline procedure should stabilize public finance while government fiscal policy is enacted. The objective is not simply to maintain zero public balance or to stabilize the level of public debt around a ceiling, but also to allow fiscal policy to proceed. In that sense, fiscal discipline may contribute to government efficiency, by forcing further rationalizations of public expenditure and revenue. Indeed, if a government is sufficiently well disciplined in its fiscal policy, it will be efficient in its use of public resources (government efficiency). The role of fiscal discipline is thus to ensure the sustainability of public finances, which implies a proper use of public resources. This first connection suggests that fiscal rules may entail greater government efficiency.

On the other hand, the literature on government efficiency aims to define what rational public expenditure and revenue may be. Reverse causality should also be considered here. Indeed, efficiency implies limited resource waste and achieving given objectives with less funding (through taxes, borrowing or the compression of productive spending). Efficiency also points to greater administrative and logistical capacities and fewer problems of coordination between budgetary units. In that sense, efficiency implies a capacity to reach aggregate targets such as those enshrined in a fiscal rule. In other words, efficient governments are less affected by the common pool problem, one of the key sources of deficit bias. Finally, government efficiency signals a stronger underlying commitment to avoiding excessive taxation of current (government size) and future generations (public debt).

3. VARIOUS FORM OF FISCAL RULES IN THE WORLD: AN OVERVIEW

Numerical fiscal rules are the simplest form of fiscal discipline measure implemented across the world. Nowadays, most countries have fiscal rules of some kind, but their details vary widely in terms of their targets, levels of application and their strictness or flexibility.

3.1 Fiscal Stance Indicators Targeted by Fiscal Rules

The variety of fiscal rules is related in part to the large number of possible targets. Numerical fiscal rules set a limit on a given public finance indicator but fiscal positions can be assessed in several ways, either through the public budget balance (or a variant such as the primary balance, cyclically adjusted balance or the balance net of public investment) or the level of public debt, public expenditure or public revenue.

A fiscal rule may first of all set a limit on the *overall budget balance*. The pros of this option are that the overall balance is a complete indicator of all the components of public balance. The cons, however, are that these include components on which the government cannot act such as the cyclical balance and components inherited from past governments. The numerical fiscal rule can also apply to the *primary balance*, defined as the overall budget balance minus components related to debt accumulation (debt service). This type of rule is less restrictive for highly indebted countries and avoids current governments having to deal with their predecessors' burdens.

An alternative target is the *cyclically adjusted balance (also called structural balance)*, which corresponds to the budget balance without cyclical components. Since governments cannot act on the cyclical part of the budgetary balance, this rule enables automatic stabilizers to operate freely. How best to calculate it is a matter of strong debate however.[2] Estimates can differ from one method to the next.

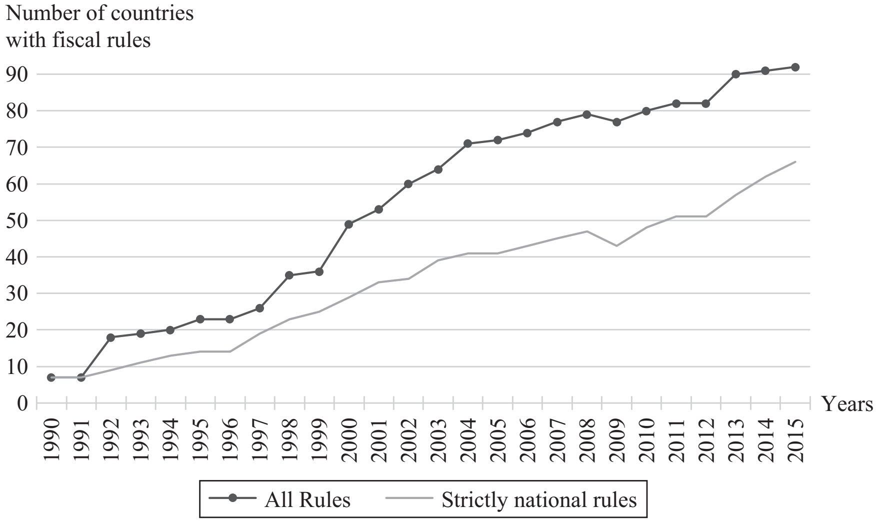
Finally, the *golden rule* (or *public investment-friendly rule*) applies to the public balance excluding public investments (meaning without public gross fixed capital formation). Under this rule, borrowing is allowed, but only to finance public investments, as a means to boost economic growth. Many authors such as Balassone and Franco (2001), Blanchard and Giavazzi (2003) and Creel (2003) have thus promoted the golden rule. Levels of public investment have indeed fallen considerably in the euro area since the launch of the single currency, even when economic conditions were not unfavorable. This would seem to support proponents of the golden rule; however, accurately assessing net public investment, as required by the golden rule, is difficult. There is no real consensus on how to assess capital depreciation and the most commonly used measure for net public investment is gross fixed capital formation. Although this approach has been used for a long time in countries such as the United Kingdom, the unresolved problem of estimating net public investment still discourages governments from adopting the golden rule.

Fiscal rules that exclude public investment or cyclical components belong to the group of more flexible or 'second-generation' fiscal rules. More elaborate 'third-generation' fiscal rules have also been proposed recently in the spirit of the Taylor rule for monetary policy, giving rise to the *fiscal Taylor rule*. Debrun and Jonung (2019) define a Taylor 'rule' as simultaneously simple, flexible and non-enforceable, and illustrate how a rules-based fiscal framework centered on a fiscal Taylor rule could work. There is no enforceable numerical rule in this case. We elaborate on the potential role for Taylor-type indicative rules in formally guiding discretion in the short run and in promoting long-term debt sustainability. In the same vein, following the series of deep economic crises since 2008 (the global financial crisis, the sovereign debt crisis and the pandemic crisis since 2020), fiscal rule compliance has been discredited (Caselli et al., 2018) and the trade-off between fiscal compliance and government growth objectives has been at the center of debates. Blanchard et al. (2021) have even proposed to abandon fiscal rules and instead adopt 'fiscal standards' defined as 'qualitative prescriptions that leave room for judgement together with a process to decide whether the standards are met'.

3.2 Numerical Fiscal Rules in the World

Fiscal rules can therefore aim at a wide variety of targets. As such, the IMF and the European Commission regularly compile data on the fiscal rules in force since 1990 in OECD countries. The dataset covers all types of numerical rules (applying to budget balance, debt, expenditure and revenue) at all levels of government (central, regional and local, general government and social security).

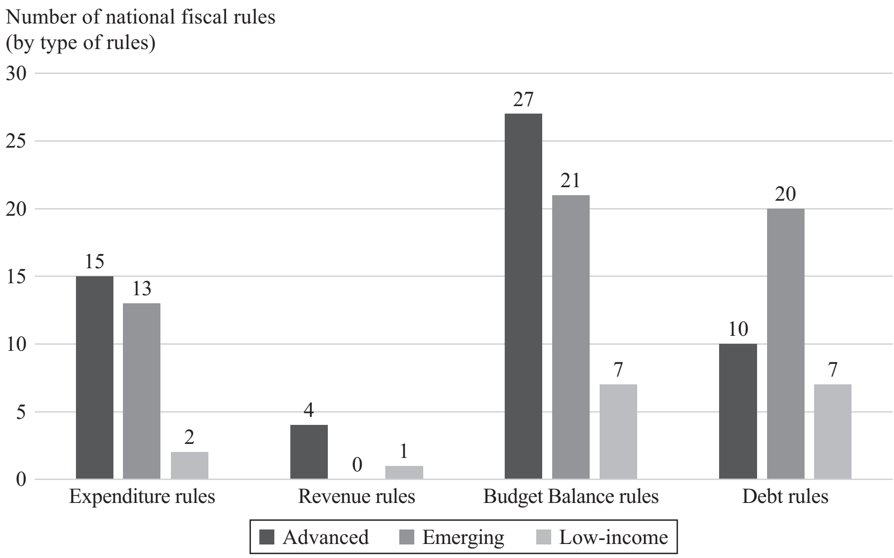
As shown in Figure 7.1, around 80 countries in the world had national and/or supranational fiscal rules in 2009, compared with just seven countries in 1990. This number has increased considerably since the 1990s with an acceleration from around 1996. This rapid expansion reflects the adoption of national fiscal rules, particularly in Europe and Latin America, or supranational rules, especially in low-income countries. In 2009, more than 50 countries had national fiscal rules (including 20 in combination with supranational rules), see Hamid et al. (2022).



Source: Authors based on IMF fiscal rules database.

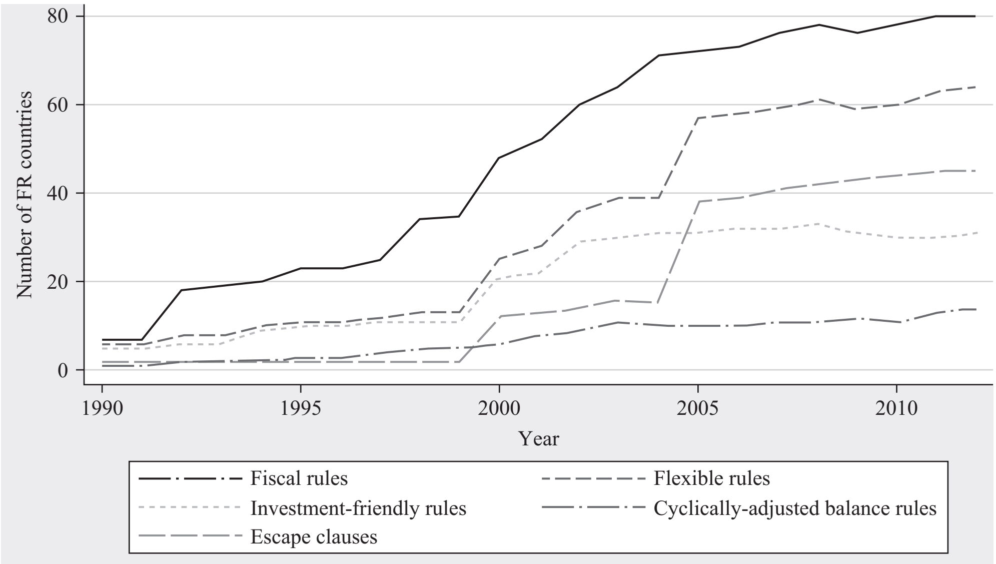
Figure 7.1 Number of countries with fiscal rules, in the world, between 1990 and 2015

As illustrated with Figure 7.2, the most popular targets of these fiscal rules in 2015 were public balance and public debt. Almost 80 percent of developed countries applied a fiscal rule based on public balance, and the same proportion a rule on public debt. This reflects a certain preference of governments for indicators closely linked to public finance sustainability.



Source: Authors based on IMF fiscal rules database.

Figure 7.2 Number of national fiscal rules by type of rule and by type of country in 2015



Source: Guerguil et al. (2017).

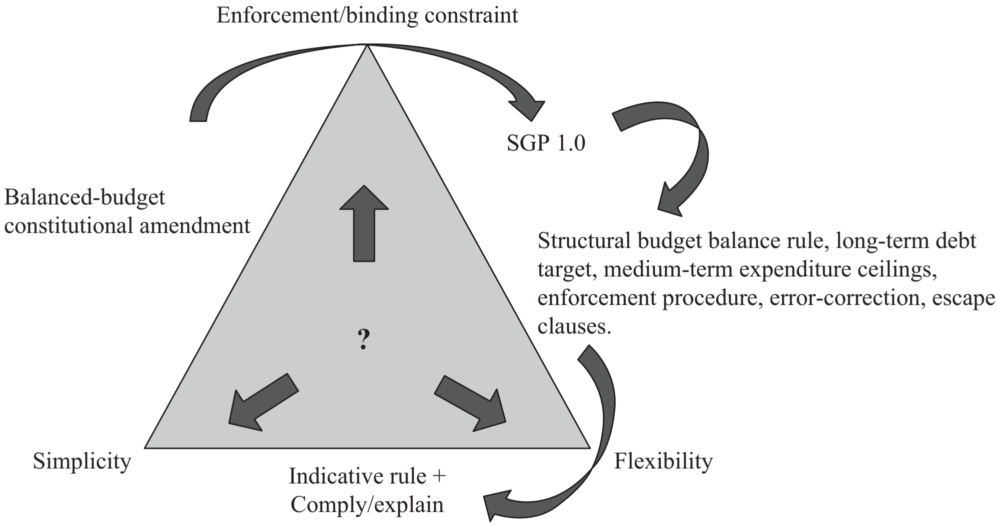
Figure 7.3 Fiscal rules adopted worldwide from 1990 to 2015

However, as highlighted by Figure 7.3, from Guerguil et al. (2017), fiscal rules have become increasingly flexible in their design over time (note that investment-friendly rules, which do not apply to public capital spending, and cyclically adjusted balance-based rules are considered flexible in this context).

4. FISCAL RULES PERFORMANCE: A BURNING ISSUE

The performance of fiscal rules is a hotly debated subject. What exactly is an efficient fiscal rule and what criteria should be used to assess its performance? Since the founding work of Kopits and Symansky (1998), many subsequent studies have attempted to deepen one or other aspect of their analysis. Although Kopits and Symansky do not explicitly refer to 'fiscal rule performance', their contribution is the origin of the debate around what constitutes an 'efficient' fiscal rule. They indeed identify eight essential properties of an 'ideal' fiscal rule, although no rule can in fact combine all eight. The optimal trade-off has to be chosen between these different qualities and several studies have attempted to do this. Thus, based on the same analysis grid, Buti et al. (2003) show that the initial supranational fiscal rule in the eurozone's Stability and Growth Pact is better than a deficit rule excluding public investment, while Creel (2003) demonstrates the exact opposite. In doing this, Creel (2003) emphasizes the subjective nature of the values associated with each property. Buiter's (2003) work is in the same vein.

More recently, Debrun and Jonung (2019) have framed the problem as the 'fiscal rule trilemma', namely that a good rule, on top of being discipline inducing and enforceable, should also be flexible (that is, sufficiently loose to not interfere too often with other policy objectives)



Source: Debrun and Jonung (2019).

Figure 7.4 Designing fiscal rules: a trilemma

and simple. However, among these three desirable properties, they explain that only two can ever be fulfilled at once. Just as in the monetary policy trilemma described by Mundell and Fleming in the 1960s, fiscal rules cannot at the same time be enforceable, flexible and simple, as illustrated in Figure 7.4.

In other words, there is no consensus on this issue. In a nutshell, the performance of fiscal rules can be understood in several non-mutually exclusive ways: i) with regard to their disciplining effect on government behavior, ii) with respect to their effect on macroeconomic variables, iii) using the fiscal rule strength index and/or iv) on the basis of how well they are complied with.

4.1 Fiscal Rule Performance as their Disciplinary Effect on Government Behavior

A first strand of the literature considers that the effectiveness of fiscal rules should be measured by their disciplinary effect on fiscal behavior, in keeping with Kopits and Symansky's (1998) first criterion: 'a good fiscal rule should be suitability for the intended objective'. Since Bohn and Inman's (1996) seminal paper, much attention has been paid to the impact of fiscal rules on the fiscal stance (assessed by an indicator of public balance such as the cyclically adjusted primary balance (CAPB) and/or public debt).[3] CAPB is commonly used as a proxy for the fiscal stance and many studies thus focus on the impact of fiscal rules on the CAPB, an increase in CAPB being interpreted as an improvement in fiscal discipline. Several studies using quite different technical approaches (instrumental variables, system Generalized Method of Moment, Least Square Dummy Variable and propensity score matching) have found that this is indeed the case. The fact that these different approaches produce the same result, as highlighted by Heinemann et al.'s (2018) metaregression analysis, is an answer of sorts to the question of possible reverse causality discussed by Debrun and Kumar (2007) and Wyplosz (2013).

Finally, while fiscal rules assessed by the CAPB appear to be ineffective in Debrun and Kumar (2007), Escolano et al. (2012) and Caselli and Reynaud (2019), many other studies conclude the opposite, such as Debrun et al. (2008), Marneffe et al. (2010), Reuter (2015) and Combes et al. (2018). These studies all focus on national fiscal rules, but positive results have also been reported for sub-national fiscal rules by Foremny (2014). The effect of fiscal rules on fiscal discipline also depends on the methodology used and the type of fiscal rule considered. While similar effects of budget balance rules on fiscal discipline have been reported by several studies (Debrun et al., 2008; Tapsoba, 2012; Barbier-Gauchard et al., 2021a. and so on), in keeping with previous studies (for example Debrun et al., 2008 and Bergman et al., 2016), they find that expenditure rules had no significant effect on the CAPB. The effect of debt rules on fiscal discipline is likewise unclear, with Barbier-Gauchard et al. (2021a) finding none, but Bergman et al. (2016) and Debrun et al. (2008) finding that debt rules do promote fiscal discipline, respectively by themselves and when combined with budget balance rules.

The performance of fiscal rules may also depend on the environment (Tapsoba, 2012; Combes et al., 2018; Reuter, 2019; Barbier-Gauchard et al., 2021a). For example, Tapsoba (2012) found that supranational rules reduce the effect of national fiscal rules in developing countries. In the EU, on the contrary, Barbier-Gauchard et al. (2021a) found no interaction between supranational and national fiscal rules when fiscal discipline is measured by the CAPB but found that there was a positive effect when fiscal discipline was proxied by a larger indicator, the global fiscal performance index (GFPI, which represents a broader definition of fiscal discipline than the CAPB does, covering the public deficit, fiscal revenue, the external deficit, the growth rate of public debt and the growth rate of public debt interest). The CAPB is indeed an imperfect measure, first, because there is no consensus on how to estimate potential production and the output gap, and second, because the CAPB only captures discretionary public spending, excluding other macroeconomic variables that may be affected by fiscal rules. This is problematic because if fiscal rules have negative effects on these other macroeconomic variables (while nevertheless reducing discretionary spending), their effectiveness is in fact limited.

4.2 Fiscal Rules' Performance as their Macroeconomic Impact

A second strand of the literature considers that the performance of fiscal rules should be judged on their impact on various macroeconomic aggregates (activity, employment, inflation, interest rates, terms of trade and so on). Fiscal rules should also be evaluated in terms of their potential stabilizing effects on these macroeconomic aggregates, in keeping with Kopits and Symansky's (1998) 'general consistency' and 'flexibility' criteria. The main outcomes considered by these studies is the volatility of one or more of these variables (typically economic activity, employment, inflation, interest rates and/or the terms of trade). For example, Sacchi and Salotti (2015) emphasize that when strict fiscal rules are introduced, discretionary policy becomes output stabilizing rather than destabilizing (more so with balanced budget rules than with expenditure, revenue or debt rules). However, fiscal rules do not alter the inflation destabilizing nature of discretionary policy, probably because of the greater role of central banks in that respect. Guerguil et al. (2017) assess the impact of different types of flexible fiscal rules on the procyclicality of fiscal policy, finding notably that investment-friendly rules reduce the procyclicality of both overall and investment spending. The effect appears to be stronger in bad times and when the rule is enacted at the national level. The existence of escape clauses in fiscal rules does not seem to affect the cyclical stance of public spending, and the inclusion of cyclical adjustment features in spending rules yields broadly similar results. Bergman and Hutchison (2015) and Combes et al. (2017) show that fiscal policy is counter-cyclical but that there are also non-linear effects. All fiscal rules do not reduce the procyclicality of fiscal policy when debt is high. Reuter et al. (2018) confirm previous findings that fiscal rules reduce fiscal volatility and consequently contribute to reducing output volatility.

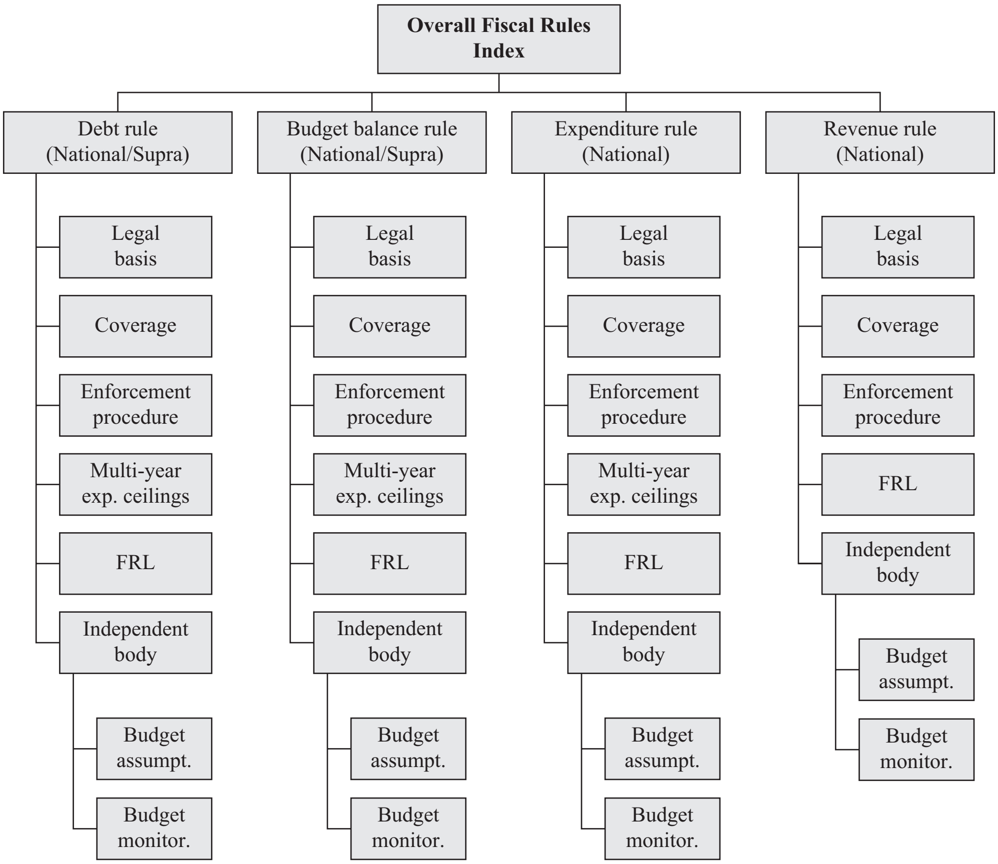
In addition, Badinger and Reuter (2017) provide strong evidence that countries with more stringent fiscal rules have higher fiscal balances (lower deficits), lower interest rate spreads on government bonds and lower output volatility. Finally, Badinger et al. (2017) and Afonso et al. (2021b) have investigated the relationship between national fiscal balances and current accounts with an emphasis on the role of fiscal rules. In agreement with previous studies, they find a positive association between fiscal and current account balances, supporting the twin deficit hypothesis. However, this effect depends on the stringency of the fiscal (budget balance or debt) rules in place.

4.3 Fiscal Rules Strength Index

It was not until the end of the 2000s that composite indicators appeared to assess the performance of fiscal rules via the fiscal rule strength index. The characteristics considered by the European Commission and the IMF in assessing the strength of a fiscal rule are its legal status, its resilience to shocks, the existence of correction mechanisms (sanctions for non-compliance in particular), the existence of institutions in charge of implementing the rule and enforcing compliance (the more independent the institution is, the better it is scored in the European Commission's evaluation grid), how easy it is to revise objectives (for example, can this be done by government alone or is parliamentary approval required?).

The European Commission has been using a fiscal rule index (FSI) and fiscal rule strength index (FRSI) since 2015 to measure the stringency of fiscal rules currently applied in the EU. This assessment is largely based on the institutional characteristics of the rules and covers at least two of Kopits and Symansky's (1998) criteria: 'flexibility' and 'credibility'. The five differently weighted criteria considered in the FRSI are i) the legal basis of the rule, ii) whether it is binding or not (can objectives be set or revised), iii) the nature and independence of control and enforcement bodies (and whether they perform real-time monitoring), iv) the correction mechanisms envisaged in case of deviation from the rule and v) resilience to shocks or events outside the control of the government. These items are scored from 0 and 1 and combined with different weights, to give the FRSI of a given fiscal rule at a given time. The FRI of each country, evaluating the overall level of fiscal discipline, is then obtained by combining the FRSIs of the different fiscal rules applied. These data are all described and accessible in the European Commission's Fiscal Governance Database.

The IMF uses a slightly different approach to calculate FRIs in its Fiscal Rules Database. Following Kumar et al. (2009), the IMF's method is based on a principal component analysis of the following sub-indices/scores: an enforcement score, a coverage score, a legal basis score, a supranational rules score, an index assessing how compliance is monitored and enforced, a flexibility score, the average number of fiscal rules and the ratio of national to total fiscal rules in each country. Schaechter et al. (2012) then developed an index based only on a set of key fiscal rule characteristics, and a series of sub-indices for each type of rule (described in Figure 7.5), with each sub-index being the sum of scores for the legal basis, coverage, formal



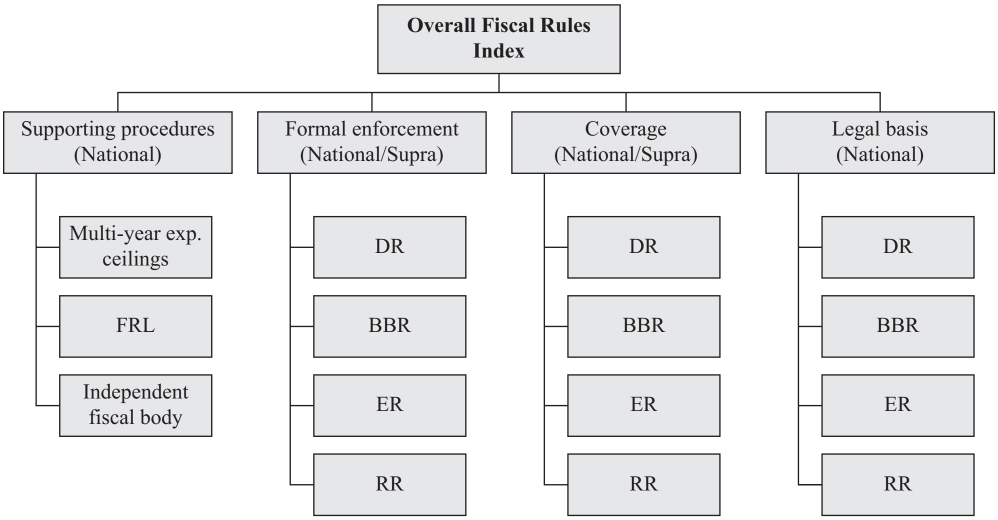
Source: IMF Staff, Schaechter et al. (2012).

Figure 7.5 IMF fiscal rule index for each type of fiscal rule

enforcement procedure and expenditure ceilings of the rule, and for the presence of a fiscal responsibility law (FRL), and of an independent body to set budget assumptions and monitor implementation (see Figure 7.6 for an overview). The four sub-indices for each type of rule are then combined to obtain the overall FRI.

4.4 Fiscal Rule Compliance Index

Among the essential properties of a 'good' fiscal rule defined by Kopits and Symansky (1998), credibility is a particularly hot topic in debates on fiscal rule efficiency. The credibility of a fiscal rule affects its enforceability (one of the elements of Debrun and Jonung's (2019) 'fiscal rules trilemma') and the level of compliance achieved. Fiscal rule compliance can be defined as how closely the fiscal aggregates considered (the budget balance, the debt-to-GDP ratio and/or government expenditure) match the targets set by the rule (without considering escape clauses).



Source: IMF Staff, Schaechter et al. (2012).

Figure 7.6 IMF fiscal rule sub-indices for each type of rule

Compliance with fiscal rules has been widely studied. Some authors have attempted to assess the level of compliance using the fiscal rules databases published by the European Commission (2022) and the IMF (see the background paper by Schaechter et al., 2012). These databases provide information on the characteristics and coverage of a given fiscal rule, its statutory base, monitoring bodies and the correction mechanisms in place in case of deviations from the rule, as well as experiences of the rule. Reuter (2015) and Larch and Santacroce (2020) show that numerical fiscal rules are generally respected in only 50 percent of cases. Delgado-Téllez et al. (2017) and Cordes et al. (2015) found similar compliance rates at the subnational level in Spain and for public expenditure rules in advance and emerging countries, respectively.

Other studies have searched for the key determinants of fiscal rule compliance. Reuter (2019), for example, did this for the European Union from 1995 to 2005, finding in particular that the characteristics of the fiscal rule (its legal basis and the existence of independent monitoring and enforcement authorities), the degree of government fragmentation and the political cycle where significantly associated with the likelihood of compliance. Compliance in these countries was not found to be affected by the existence of overlapping fiscal rules (at the national level, or with regional or local rules). Larch and Santacroce (2020) studied the determinants of compliance with the supranational fiscal rules introduced by the European Union's Stability and Growth Pact (which was adopted in 1997). These rules, which have been reformed several times, set targets for several fiscal aggregates (deficit, debt, structural balance and expenditure). These authors' study covering 1998 to 2019 highlights persistent differences between countries. Numerical compliance seems to be linked with key macro-economic variables on the one hand (especially episodes of pro-cyclical fiscal policy) and the quality of governing institutions on the other (countries with 'watchdogs' (Debrun and Jonung,2019), that is national independent fiscal institutions tend to be more compliant).

Elsewhere, Barbier-Gauchard et al. (2021b) used a machine learning based forecasting model to predict compliance with the 3 percent deficit rule. The dataset covered the years 2006 to 2018 (a turbulent period that included the global financial crisis and the sovereign debt crisis) for all 28 EU member states. Eight of the 141 variables considered were identified as predictors via a feature selection procedure and the proposed model had a forecasting accuracy of 91.7 percent, outperforming the logit model that was used as a benchmark.

5. AN ASSESSMENT OF THE RELATIONSHIP BETWEEN GOVERNMENT EFFICIENCY AND FISCAL RULES

This large literature on fiscal rules does not contain any empirical evidence of a potential link between fiscal rules and government efficiency. This leaves several questions outstanding. Do fiscal rules impact government efficiency? Or do governments adopt fiscal rules in response to poor government efficiency? Are efficient governments more likely to use fiscal rules than others or less likely? There are several issues to investigate and this section explores several aspects of the relationship between fiscal rules and government efficiency. Using fiscal rule related proxies, we will test the links between government efficiency and i) the existence, ii) the stringency and iii) the level of compliance with fiscal rules.

Government efficiency and its determinants have been studied by Gupta and Verhoeven (2001) and Afonso et al. (2005), who measured government efficiency through the effect of public expenditure on socio-economic indicators. Both sets of authors used a non-parametric production function approach and obtained efficiency scores from the distances between the inefficient observations and the frontier. More recently, Afonso et al. (2021b) calculated a novel measure of government efficiency for 36 advanced OECD economies over the period 2003–17. Their government efficiency indicator was calculated in five-year periods: 2003–07, 2008–12 and 2013–17. Table 7.2 provides descriptive statistics on this indicator for the 34 advanced OECD countries considered in this chapter,[4] with details on the fiscal rules considered given in Appendix A1. The government efficiency index varies little on average, the mean index ranging from 0.95 to 1.17 across the three studied periods. Nevertheless, the extrema (the minimum and the maximum of the sample) vary considerably. It may therefore be interesting to investigate whether fiscal rules played a role in extreme cases, in other words, whether they helped in countries with particularly efficient or inefficient governments.

Table 7.2 Descriptive statistics on government efficiency for 34 advanced OECD economies

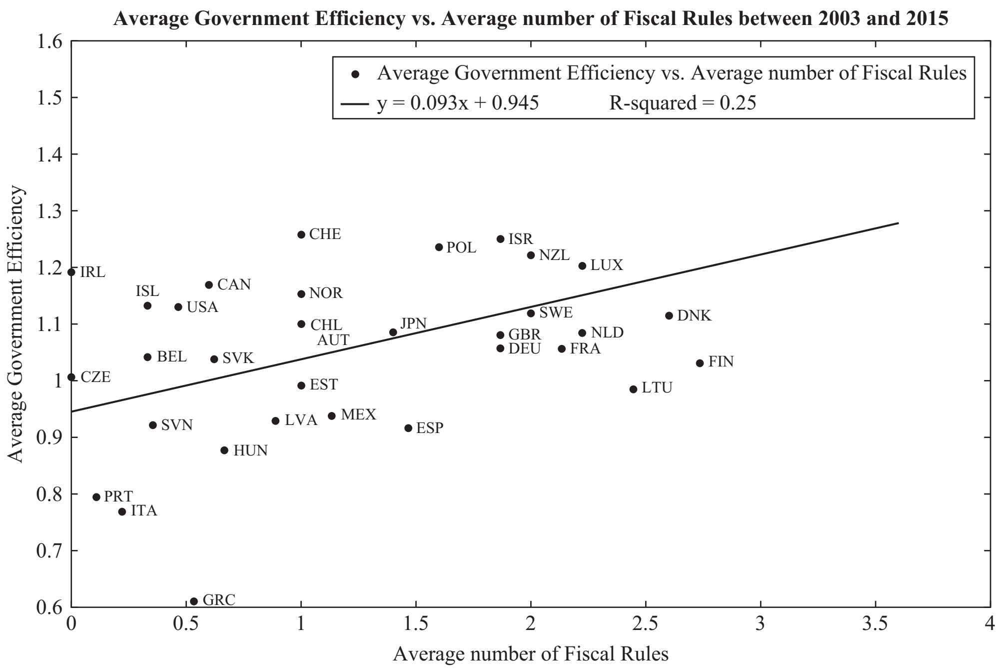
Variable	Period	min	max	Mean	Median
Government Efficiency	2003–07	0.952	1.375	1.170	1.186
Government Efficiency	2008–12	0.014	2.056	0.948	0.956
Government Efficiency	2013–15	0.672	1.581	1.066	1.032

Source: Afonso et al. (2021b).

5.1 Government Efficiency and the Presence of Fiscal Rules

Government efficiency is measured over five-year periods[5] while the presence of fiscal rules is a year-to-year measure. To identify a potential link between the two, we calculate the annual average of the number of fiscal rules in place in each country over each five-year period, considering fiscal rules (budget balance rules, expenditure rules, revenue rules and debt rules) together and separately. In other words, we count how many fiscal rules a given country had on average, each year, over the considered period.[6] This provides a time-invariant measure for fiscal rules.[7] This approach is used for the overall number of fiscal rules and for each type of rule separately in the following paragraphs. The following figures reflect unconditional correlations and do not imply causality, hence fitted lines should be interpreted with caution.

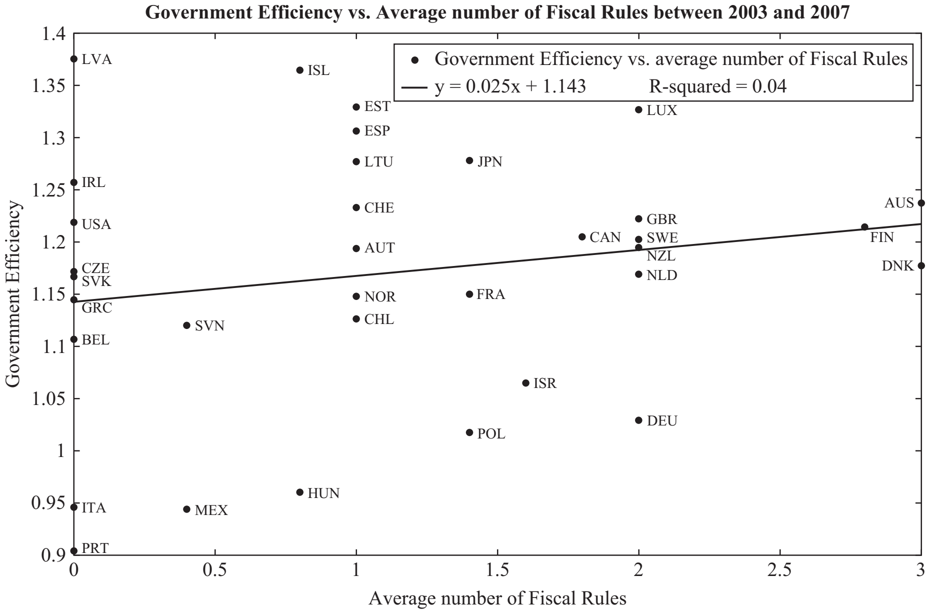
For *all national fiscal rules*, Figure 7.7.a plots government efficiency on the vertical axis and the annual average of the number of fiscal rules on the horizontal axis for the entire 2003–15 period. The two quantities appear to be positively related but we know that in this period, two economic crises (the global financial crisis and the sovereign debt crisis) affected countries worldwide. It may therefore be informative to decompose the study period into three parts (2003–07, 2008–12, 2013–15[8]). The results of this decomposition are shown in Figure 7.7.b.

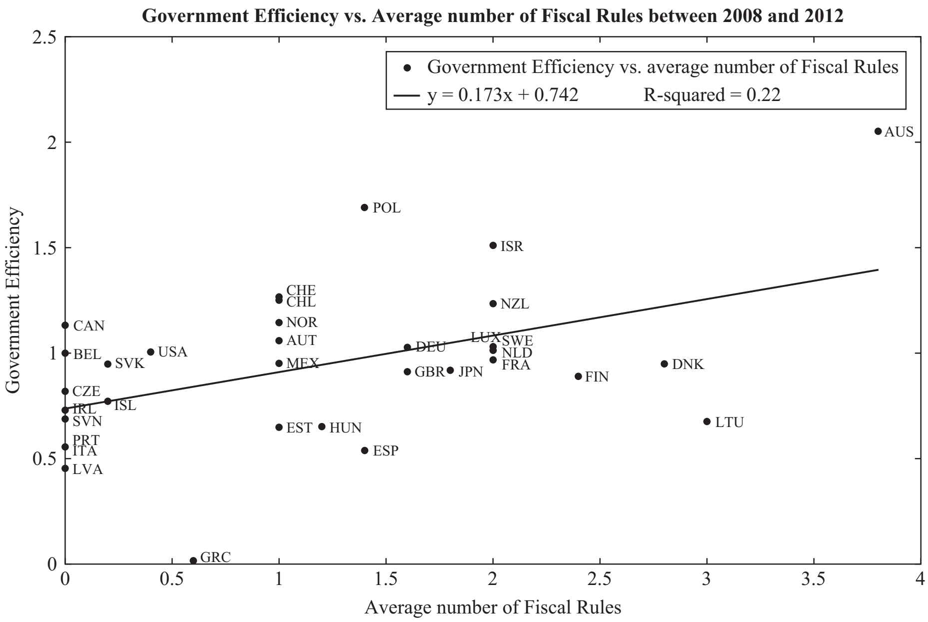


Note: The values represented in this graphic are the annual averages of government efficiency and of the number of fiscal rules calculated across the three periods (2003–07; 2008–12; 2013–15).

Source: Authors based on Government Efficiency index from Afonso et al. (2021b) and IMF Fiscal rules database.

Figure 7.7.a Government efficiency and the annual average number of fiscal rules from 2003 to 2015





Note: The x-axis values are averages over the considered period of the number of national fiscal rules in place in a given country in a given year. For example, an average number of fiscal rules equal to 1 means that, on average, the country had one fiscal rule in place in each year of the considered period. Linear regression was used for smoothing.

Source: Authors based on Government Efficiency index from Afonso et al. (2021b) and IMF Fiscal rules database.

Figure 7.7.b Government efficiency and the average number of fiscal rules by period

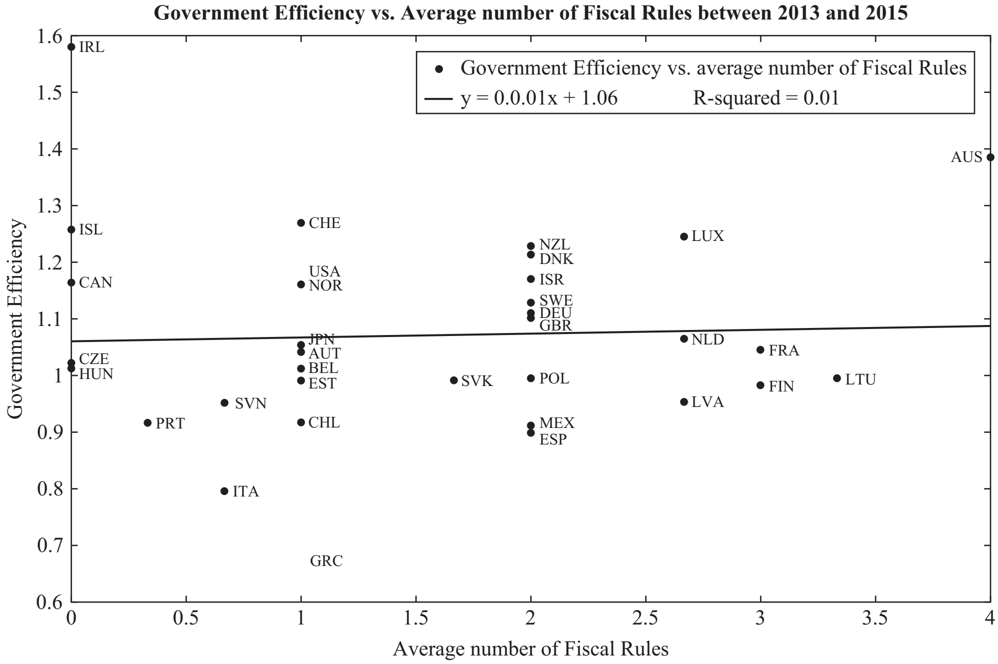


Figure 7.7.b (Continued)

The correlation coefficients between the annual average of the number of fiscal rules and government efficiency for the three periods (listed in Appendix A2) support the visual impression provided by Figure 7.7.b. The presence of fiscal rules is weakly correlated with government efficiency between 2003 and 2007 but more strongly correlated between 2008 and 2012. There are several possible explanations for this. First, many fiscal rules were broken during the great recession, because governments were no longer constrained and used public investment to drive economic recovery, a policy that can reasonably be considered efficient. Second, governments that were constrained by fiscal rules may have shifted from bad spending to highly productive investments. However, since all governments were forced to use fiscal stimulus in the face of the crisis, it is unclear whether fiscal rules had anything to do with these decisions. Overall, the data suggest that fiscal rules have a weak positive effect on government efficiency, but a deeper analysis suggests that the results depend on the period considered. Another potential source of heterogeneity in the results is the variety of fiscal rules applied. We will now investigate this link separately for each type of rule.

To further explore the potential links between fiscal rules and government efficiency we now consider the strength of fiscal rules in place rather than simply their number.

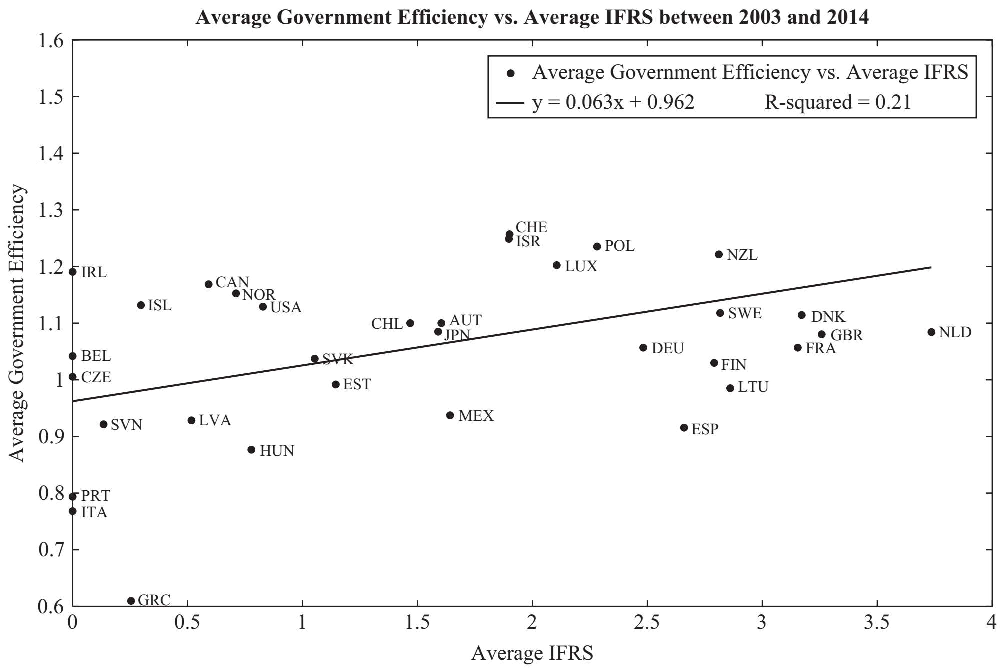
5.2 Government Efficiency and the Stringency of Fiscal Rules

As for the number of fiscal rules, we transformed the annual measure of fiscal rule stringency to obtain a variable that can be compared with government efficiency. We calculated the mean index of fiscal rule strength (IFRS) over each period 2003–07 and 2008–12.[9] The IFRS values came from Schaechter et al. (2012) and Combes et al. (2018).

Figure 7.8.a suggests that government efficiency is positively related to the average IFRS: the stronger the rules in place are, the more efficient governments seem to be. Considering the two periods 2003–07 and 2008–12 separately, Figure 7.8.b shows that the correlation is stronger in the latter period than in the former. In contrast with previous results therefore, these data suggest that strong rules do alter trends in public spending.

5.3 Government Efficiency and Fiscal Rule Compliance

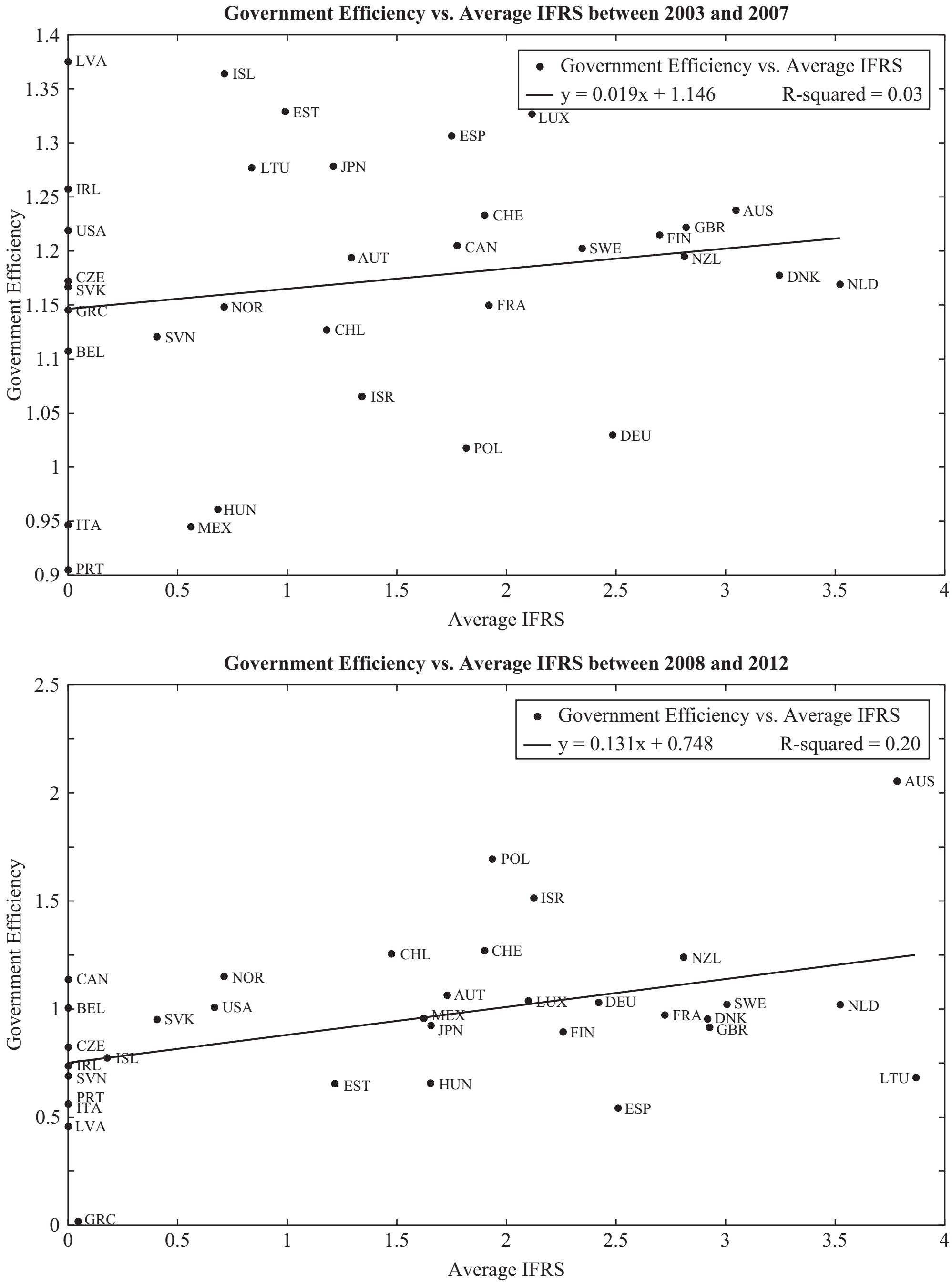
The analysis for this section is based on Reuter's (2015) definition of compliance. Fiscal rules compliance's computations used a dataset shared by Reuter according to the work done in Caselli et al. (2018), completed by our own calculations of compliance with national fiscal rules. To explore the link between government effectiveness (not efficiency) and average



Note: Government efficiency was considered constant between 2013 and 2017, its average is thus equal for 2013–14 and for the 2013–17 period. Since IFRS values were only available until 2014, we calculated average values over three periods (2003–07; 2008–12; 2013–14).

Source: Authors based on Government Efficiency index from Afonso et al. (2021b) and IFRS from Schaechter et al. (2012) and Combes et al. (2018).

Figure 7.8.a Government efficiency and index of fiscal rule strength from 2003 to 2014



Source: Authors based on Government Efficiency index from Afonso et al. (2021b) and IFRS from Schaechter et al. (2012) and Combes et al. (2018).

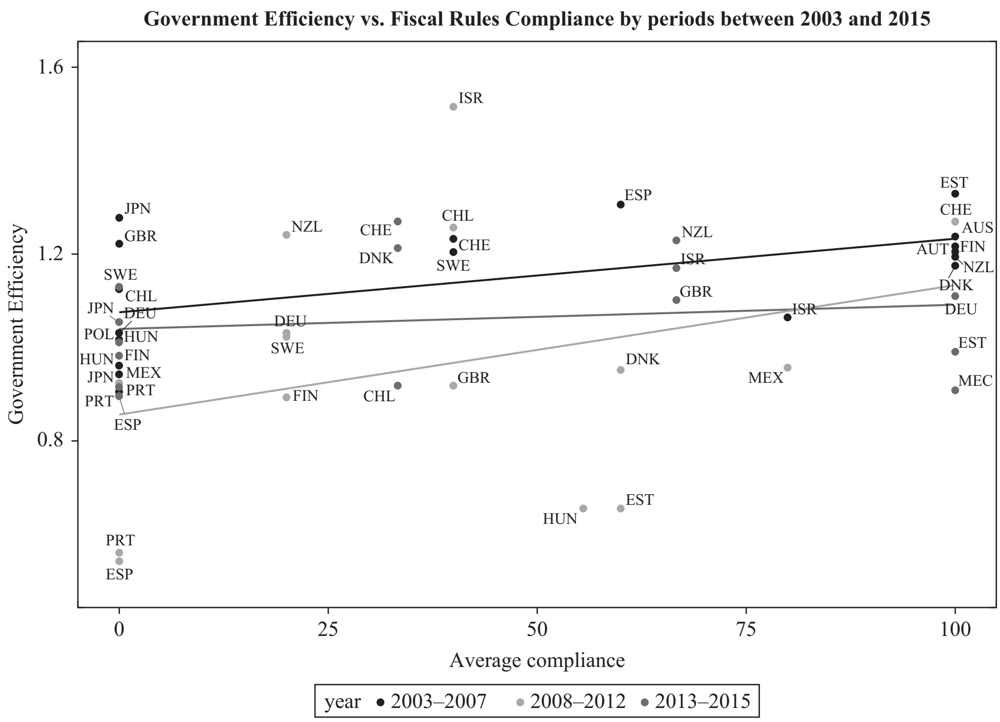
Figure 7.8.b Government efficiency and index of fiscal rule strength by period

compliance with supranational rules, readers are referred to Larch and Santacroce (2020), who focus on the Stability and Growth Pact's rules. Following these authors, we calculate the average fiscal rule compliance score, C_i, for each country i, over the three considered periods (2003–07; 2008–12; 2013–15) as follows:

$$C_i = \frac{\sum_t \sum_r C_{i,r,t}}{\Sigma_t \Sigma_r} \tag{7.1}$$

The analysis is based on average rates over countries (i), time (t) and rules (r). Averages are calculated by counting instances of compliance in a given period and dividing by the total number of rules to be complied with. Considering for example a country with only one type of fiscal rule (say a Budget Balance Rule) applied every year from 2003 to 2007, but complied with in only three of the five years, the average compliance score is 60 percent. Now, if two fiscal rules are applied in the period, the first one only for three years and complied with once, and second one in place for five years and complied with twice, the average compliance score is 37.5 percent.

Of the 34 countries included in the preceding analyses in this chapter, we only consider 19 here. This is because Reuter's compliance database refers mainly to budget balance rules. While we were able to complete the database in some cases with our own calculations, several



Source: Authors based on Government Efficiency index from Afonso et al. (2021b) and average compliance index from Caselli et al. (2018) and authors' calculations.

Figure 7.9 Government efficiency and fiscal rule compliance by period countries in our initial set of 34 do not have a budget balance rule. We thus calculated the association between average compliance with BBRs and government efficiency. Scores were obtained for each country in each period during which a BBR was in place. These results are shown in Figure 7.9 for 2003–07, 2008–12 and 2013–15.

Figure 7.9 shows that the correlation between fiscal rule compliance and government efficiency was weaker (the regression line is less steep) in the 2003–07 period, in the so-called pre-global financial crisis (GFC) period. This reflects a relaxation of expenditure controls other than compliance with fiscal rules. The link seems to disappear after the sovereign debt crisis (the slope of the regression line is close to zero), between 2013 and 2015. However, the link between fiscal rule compliance and government efficiency seems to be strongest (the regression line is steepest) between 2008 and 2012, suggesting that countries that complied with rules during the GFC invested more efficiently. The weak association between fiscal rule compliance and government efficiency before and after the GFC, suggests that the increased correlation during the GFC may be driven by governments' implementation of fiscal plans in response to the crisis. The corresponding increase in government efficiency is not causally related to fiscal rule compliance. This interpretation is consistent with the low level of compliance in this period, where escape clauses were activated, such that compliance is unlikely to have had any impact during the crisis.

Table 7.3 lists the main findings of this exploratory work. There appears to be a link between fiscal discipline and government efficiency, whether fiscal discipline is assessed by

Table 7.3 Sign of the relationship between fiscal discipline and government efficiency for different indicators of fiscal discipline from 2003 to 2015 (by period and sub-period)

Indicator	Existence (all rules)	Stringency*	Compliance
Periods	2003–15	2003–14	2003–15
Intensity of the relation between fiscal discipline and government efficiency	++	++	+

Indicator	Existence (all rules)			Stringency*			Compliance		
Sub-periods	2003–07	2008–12	2013–15	2003–07	2008–12	2013–15	2003–07	2008–12	2013–15
Intensity of the relation between fiscal discipline and government efficiency	+	++	null	+	++	not available	+	++	null

Note: *data only between 2003 and 2014.

Source: Authors' calculations.

the existence of fiscal rules, their strength index or their compliance rate for the 34 countries in the study from 2003 to 2015. Analyses by sub-periods provide additional insights into the cyclical nature of this association.

6. CONCLUSIONS AND AVENUES FOR FUTURE RESEARCH

This chapter provides an overview of fiscal performance, which can be measured by fiscal effectiveness or fiscal efficiency. The two definitions are not equivalent and should be discussed with caution. The link between fiscal rules and government efficiency has already been well studied in the literature, including through many empirical investigations. Although some studies have found otherwise, the general consensus is that fiscal rules have a positive effect on fiscal discipline and fiscal effectiveness.

This chapter offers some insight into the potential link between fiscal rules and government efficiency by investigating the correlation between the presence, strength and compliance rate of fiscal rules and a government efficiency index. The results do not support the existence of a strong relationship between fiscal rules and fiscal efficiency. The outcomes of the analyses depend on the indicator used to proxy fiscal rule performance, the period considered, and the types of fiscal rule included. The simple presence of fiscal rules seems to have had a positive impact on fiscal efficiency overall between 2003 and 2015 but the effect varies over time. The link between the stringency of fiscal rules and government efficiency seems to be stronger, with stricter fiscal rules being associated with higher government efficiency. Similar results are obtained for fiscal rule compliance but these vary between time periods: compliance with fiscal rules is weakly correlated with government efficiency before, and after the global financial crisis is highly correlated during the crisis. This may reflect the activation of release clauses and the low level of compliance in the crisis, as exemplified by the United Kingdom abandoning the golden rule in 2009.

In summary, the heterogeneity of the results and the weakness of the correlation coefficients obtained are inconsistent with a strong and obvious link. Nevertheless, these findings should be interpreted with caution. First, our analysis covers a mixed set of countries. Opposite trends may thus cancel out, such that no average effect is observed. At least three scenarios can be envisaged that would contribute to this effect:

(1) In a country that is already rigorous and has sustainable public finances because of good fiscal policies, the adoption of and compliance with fiscal rules will not lead to any change since the government was already efficient in their absence.
(2) The opposite may also occur if countries adopt fiscal rules but choose to disregard their target and therefore fail to comply. Japan, for example, has never complied with its budget balance rule. The government's behavior regarding fiscal discipline cannot be captured in such cases and fiscal rules are ineffective because the government may choose to temporarily ignore its commitment to fiscal discipline. Consequently, fiscal rules may not affect fiscal efficiency, which is poor due to government choices, despite the presence of fiscal rules.
(3) We only considered national fiscal rules but different levels could also have been considered. It is possible that subnational and supranational fiscal rules induce other effects

than those observed here. For example, Sow and Razafimahefa (2015) found a non-linear relationship between expenditure decentralization and fiscal efficiency. Sub-national fiscal rules may thus affect government efficiency since they may constrain sub-national public expenditure. Another possibility would have been to use Larch and Santacroce's (2020) database, which covers supranational rule compliance and could be used to analyze the link between supranational fiscal rule compliance and government efficiency in EU economies.

These considerations lead us to recommend case studies of panels of countries with similar fiscal behaviors. This should ensure the true effects of fiscal rules are captured, independently of the fiscal context and government attitudes to their fiscal policy commitments.

Another valuable approach may be to investigate specific cases involving a strong change in fiscal behavior. The government efficiency index covers five-year periods so single-year changes in fiscal behavior cannot be observed, but an important change in government efficiency index between two periods suggests that strong reforms were implemented leading to changes in fiscal behavior from one period to the next. If a fiscal rule has been adopted between these periods, these cases would be good candidates for empirical analysis, which should reveal whether and how the fiscal rule contributed to the change in government efficiency. Robustness checks could be conducted to observe if compliance with the rule or particular characteristics such as the fiscal rule's strength increased its putative effect.

Finally, while the effect of fiscal rules on fiscal effectiveness appears to be positive on average and is easy to observe empirically, the relationship between fiscal rules and government efficiency is not obvious. There is a large avenue for future research involving a careful selection of countries based on the relevance of their fiscal practices.

APPENDICES

Table 7.A1 National rules included in the analysis

Country	Level of public finance	Constrained Public Finance Indicator	Targeted Value	From	To	Additionnal details
Australia	GG	Δ BB	0	1985	1988	
	GG	SB	0	1998		Structural balance
Canada	CG	BB	0	1998	2005	
Chile	CG	SB	1	2001	2007	
	CG	SB	[0.5; −2]	2008		Target in a range
Denmark	GG	SB	−0.5	1992	2000	
	GG	SB	0	2001	2013	
Estonia	GG	BB	0	1993	2011	
	GG	SB	0	2012		
Finland	CG	SB	[0;1]	1999	2013	1% between 2007–11
	CG	BB	[−2.75; −2.5]	1999	2008	−2.75% between 1999–2002
	CG	BB	−1	2011		
Germany	CG	OP	0	1969	2010	
	CG	SB	−0.35	2011		
Hungary	GG	PB	0	2004	2009	
	GG	Δ PB	0	2010	2011	
	CG	PB	0	2009	2011	
Israel						
Japan	CG	OP	0	1990	2015	
Mexico	CG	BB	0	2006	2008	
	CG	OP	[−1.1;0]	2009		
New Zealand	GG	OP	0	1994		
Poland	GG	BB	[−0.09; -0.11]	2006	2007	Limit of 30 Bio. PLN
Portugal	CG	BB	0	2002		
Spain	GG	BB	[0;-2]	2003	2011	Limit depending on GDP growth
Sweden	GG	BB	[1;2]	2000		1% since 2007
Switzerland	CG	SB	0	2003		

Table 7.A1 (Continued)

Country	Level of public finance	Constrained Public Finance Indicator	Targeted Value	From	To	Additionnal details
United Kingdom	GG	OP	0	1997	2008	cyclically adjusted Operating Balance
	GG	Δ BB	0	2010		

Notes: Δ means 'annual change'. BB means nominal budget balance. PB means primary balance. SB means structural balance. OP means operating balance. GG means General Government. CG means Central Government.

Source: Lledó et al. (2017), Reuter (2019), including authors' adjustments.

Table 7.A2 Correlation between the annual average of the number of fiscal rules and government efficiency

Period	Correlation
2003–07	0.19
2008–12	0.47
2013–15	0.05

Source: Authors' calculations.

NOTES

1. See Kaufman et al. (2010) for additional information on the methodology and analytical issues.
2. One possible method is to estimate the structural balance by removing all cyclical components from the overall balance (Fatás and Mihov, 2003, 2006). This residual method considers the structural public balance as the residual of the public budget after checking the position of the cycle and excluding one-off elements. Although the concept of this approach, eliminating factors considered temporary, is clear, the method can be biased because the structural public balance is prone to measurement error and 'noise', and therefore varies depending on the calculation method used. Even the most robust econometric estimators cannot permanently contain volatility (Andersen, 2013). Moreover, this method has a naïve bias because other variables such as inflation are disregarded. Another technical problem is that estimates of cyclical components require knowledge of the output gap (the difference between production and potential production). There is no consensus on the best approach to achieve this: traditional econometric filters such as Hodrick Prescott filters or the production function approach based on GDP, which uses full employment as the potential reference GDP (see Bouthevillain et al., 2001).
3. The effectiveness of fiscal rules could also be measured by their effect on electoral cycles, see for instance Gootjes et al. (2019) or Eklou and Joanis (2019).
4. The 34 countries are: Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, the United Kingdom and the United States. South Korea and Turkey are not included because their data are not included in the IMF Fiscal rules database from which we retrieve fiscal rules data.
5. This index is time invariant by period.
6. For example: a country that had one expenditure rule (ER) in place every year in the range 2003–07 has an annual average number of ERs equal to 1 for that period. A country that had one ER in place for 3 years between 2003 and 2007 has an annual average number of ERs of 0.6 for that period. Annual averages are calculated in the same way for each type of fiscal rule (expenditure rules, revenue rules, budget balance rules and debt rules), and the sum of all annual averages gives the total average number of fiscal rules.

7. We do not simply add up the numbers of each type of fiscal rule in place over a period since this could lead to double counting. Indeed, a country with one ER in place every year from 2003 to 2007 may have applied the same ER over the entire period. If we report five ERs for this period, this is more a measure of the number of years covered by an ER than the number of ERs in place over this period. On the contrary, for a country with two different ERs in place every year between 2003 and 2007, the annual average number of ERs will be two.
8. Government efficiency is calculated over 2003–07, 2008–12 and 2013–17 but since the IMF fiscal rules database (2016) only runs until 2015, we cut the last period short.
9. IFRS are available until 2014, we thus did not study the 2013–14 period.

REFERENCES

Afonso, A., Huart, F., Tovar Jalles, J. and Stanek, P. (2021a), 'Fiscal rules and twin deficits: The link between fiscal and external balances', *Journal of International Money and Finance*, forthcoming.

Afonso, A., Tovar Jalles, J. and Venâncio, A. (2021b), 'Taxation and public spending efficiency: An international comparison', *Comparative Economic Studies*, 63, 356–383.

Afonso, A., Schuknecht, L. and Tanzi, V. (2005), 'Public Sector Efficiency: An International Comparison', *Public Choice*, 123 (3–4), 321–347.

Alam, R., Kitenge, E. and Bedane, B. (2017), 'Government effectiveness and economic growth', *Economics Bulletin*, 37 (1), 222–227.

Andersen, T.M. (2013), 'Fiscal policy targeting under imperfect information', *Journal of International Money and Finance*, 34, 114–130.

Artis, M. and Onorante, L. (2006), 'The economic importance of fiscal rules', *CEPR Discussion Paper*, No. 5684.

Badinger, H., Fichet de Clairfontaine, A. and Reuter, W.H. (2017), 'Fiscal rules and twin deficits: The link between fiscal and external balances', *The World Economy*, 40, 21–35.

Badinger, H. and Reuter, W.H. (2017), 'The case for fiscal rule', *Economic Modelling*, 60, 334–343.

Balassone, F. and Franco, D. (2001), *'EMU fiscal rules: a new answer to an old question?'* In: Banca d'Italia (ed.) Fiscal rules. Rome.

Barbier-Gauchard, A., Baret, K. and Minea, A. (2021a), 'National fiscal rules adoption and fiscal discipline in the European Union', *Applied Economics*, 53, (20), 2337–2359.

Barbier-Gauchard, A., Baret, K. and Papadimitriou, T. (2021b), 'Forecasting the stability and growth pact compliance using machine learning', *BETA Working Paper*, No. 2021-01.

Bergman, M. and Hutchison, M.M. (2015), 'Economic stabilization in the post-crisis world: Are fiscal rules the answer?', *Journal of International Money and Finance*, 52, 82–101.

Bergman, M., Hougaard Jensen, S.E. and Hutchison, M.M. (2016), 'Promoting sustainable public finances in the European Union: The role of fiscal rules and government efficiency', *European Journal of Political Economy*, 44, 1–19.

Blanchard, O. and Giavazzi, F. (2003), 'Improving the SGP through a proper accounting of public investment', *CEPR European Economic Perspectives newsletter*, 1.

Blanchard, O., Leandro, A. and Zettelmeyer, J. (2021), 'Redesigning EU fiscal rules: From rules to standards', *Economic Policy*, 36 (106), 195–236.

Bohn, H. and Inman, R. (1996), 'Balanced budget rules and public deficits: Evidence from the U.S. states', *NBER Working Paper*, No. 5533.

Bouthevillain, C., Cour-Thimann, P., Van Den Dool, G., Hernández De Cos, H., Langenus, G., Mohr, M., Momigliano, S. and Tujula, M. (2001), 'Cyclically adjusted budget balances: An alternative approach', *ECB Working Paper*, No.77.

Bova, E., Carcenac, N. and Guerguil, M. (2014), 'Fiscal rules and the procyclicality of fiscal policy in the developing world', *IMF Working Paper*, WP/14/122.

Buiter, W. (2003), 'Ten commandments for a fiscal rule in the E(M)U', *Oxford Review of Economic Policy*, 19 (1), 84–99.

Buti, M., Eijffinger, S. and Franco, D. (2003), 'Revisiting the stability and growth pact: Grand design or internal adjustment?', *European Economy-Economic Papers*, No. 180.

Caselli, F. and Reynaud, J. (2019), 'Do fiscal rules cause better fiscal balances? A new instrumental variable strategy', *IMF working paper*, WP/19/49.

Caselli, F., Eyraud, L., Hodge, A., Diaz Kalan, F., Kim, Y., Lledó, V., Mbaye, S., Popescu, A., Reuter, W. H., Reynaud, J., Ture, E. and Wingender, P. (2018), 'Second-generation fiscal rules: Balancing simplicity, flexibility, and enforceability', *Technical Background Papers. Staff Discussion Notes*, No. 18/04.

Castro, V. (2011), 'The impact of the European Union fiscal rules on economic growth', *Journal of Macroeconomics*, 33, 133–326.

Combes, J-L., Minea, A. and Sow, M. (2017), 'Is fiscal policy always counter- (pro-) cyclical? The role of public debt and fiscal rules', *Economic Modelling*, 65, 138–146.

Combes, J.-L., Minea, A., Debrun, X. and Tapsoba, R. (2018), 'Inflation targeting, fiscal rules and the policy mix: Cross-effects and interactions', *The Economic Journal*, 128 (615), 2755–2784.

Cordes,T., Kinda, T., Muthoora, P. and Weber, A. (2015), 'Expenditure rules: Effective tools for sound fiscal policy?', *IMF Working Paper*, No. 15/29.

Creel, J. (2003), 'Ranking fiscal policy rules: The golden rule of public finance vs. the stability and growth pact', *Document de travail de l'OFCE*, 2003–2004.

De Grauwe, P. (2000), 'Monetary policies in the presence of asymmetries', *CEPR discussion paper*, No. 2393.

Debrun, X. and Jonung, L. (2019), 'Under threat: Rules-based fiscal policy and how to preserve it', *European Journal of Political Economy*, 57, 142–157.

Debrun, X. and Kumar, M.S. (2007), 'The discipline-enhancing role of fiscal institutions: Theory and empirical evidence', *IMF Working Paper*, No. 07/171.

Debrun, X., Moulin, L., Turrini, A., Ayuso i casals, J. and Kumar, M.S. (2008), 'Tied to the Mast? National fiscal rules in the European Union', *Economic Policy*, 23, 297–362.

Delgado-Téllez, M., Lled´o, V.D. and Pérez, J.J. (2017), 'On the determinants of fiscal non-compliance: An empirical analysis of Spain's regions.', *IMF Working Paper*, No. 17/5.

Eklou, K.M. and Joanis, M. (2019), 'Do fiscal rules cause fiscal discipline over the electoral cycle?' *IMF working paper*, WP/19/291.

Escolano, J., Eyraud, L., Sarnes, J., Moreno Badia, M. and Tuladhar, A. (2012), 'Fiscal performance, institutional design and decentralization in European Union countries', *IMF Working Paper*, No. 12/45.

Fatás, A. and Mihov, I. (2003), 'The case for restricting fiscal policy discretion', *Quaterly Journal of Economics*, 118, 1419–1447.

Fatás, A. and Mihov, I. (2006), 'The macroeconomic effects of fiscal rules in the US states', *Journal of Public Economics*, 90, 101–117.

Foremny, D. (2014), 'Sub-national deficits in European Countries: The impact of fiscal rules and tax autonomy', *European Journal of Political Economy*, 34,86–110.

Gali, J., and Perotti, R., (2003), 'Fiscal policy and monetary integration in Europe', *NBER Working Paper*, No. 9773.

García-Sánchez I. M., Cuadrado-Ballesteros B. and Frias-Aceituno J., (2013), 'Determinants of Government Effectiveness', *International Journal of Public Administration*, 36, 567–577.

Ghosh, A., Kim, J. Mendoza, E., Ostry, J. and Qureshi, M. (2011), 'Fiscal fatigue, fiscal space and debt sustainability in advanced economies', *NBER Working Paper*, No.16782.

Gootjes, B., de Haan, J. and Jong-A-Pin, R. (2019), 'Do fiscal rules constrain political budget cycles?', *DNB Working Paper*, No. 634.

Guerguil, M., Mandon, P. and Tapsoba, R. (2017), 'Flexible fiscal rules and countercyclical fiscal policy', *Journal of Macroeconomics*, 52, 189–220.

Gupta, S. and Verhoeven, M. (2001), 'The efficiency of government expenditure—experiences from Africa.' *Journal of Policy Modelling* 23, 433–467

Hamid D., Elger P., Fotiou A., Garcia-Macia D., Lagerborg A., Lam R. and Pillai S. (2022). 'Fiscal Rules Dataset: 1985-2021', International Monetary Fund, Washington, D.C.

Heinemann, F., Moessinger, M-D. and Yeter, M. (2018), 'Do fiscal rules constrain fiscal policy? A metaregression-analysis', *European Journal of Political Economy*, 51, 69–92.

IMF (2009), *Fiscal Rules – Anchoring Expectations for Sustainable Public Finances*, Fiscal Affairs Department.

Lledó, V. and Yoon, S., Fang, X., Mbaye, S. and Kim, Y. (2017), 'Fiscal rules at a Glance', *IMF Background Paper*.

Kell, M. (2001), 'An assessment of fiscal rules in the United Kingdom', *IMF Working Papers*, No. 01/91

Kopits, G. and Symansky, S.A. (1998), 'Fiscal policy rules', *IMF Occasional Paper*, No. 162.

Kopits, G. (2004), 'Overview of Fiscal Policy Rules in Emerging Markets' in G. Kopits (ed.), *Rules-based Fiscal Policy in Emerging Markets: Background, Analysis and Prospects*, Palgrave (Macmillan), Houndmills, United Kingdom

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2010), 'The worldwide governance indicators: Methodology and analytical issues', *World Bank Policy Research Working Paper*, No. 5430

Kumar, M., Baldacci, E., Schaechter, A., Caceres, C., Kim, D., Debrun, X., Escolano, J., Jonas, J., Karam, P., Yakadina, I. and Zymek, R. (2009), 'Fiscal rules—Anchoring expectations for sustainable public finances', *IMF Staff Paper*.

Kydland, F. and Prescott, E. (1977), 'Rules rather than discretion: The inconsistency of optimal plans', *Journal of Political Economy*, 85 (3), 473–491

Larch, M. and Santacroce, S. (2020), 'Numerical compliance with EU fiscal rules: The compliance database of the Secretariat of the European Fiscal Board (Database)'.

Marneffe, W., Van Aarle, B., Van Der Wielen,W. and Vereeck, L. (2010), 'The impact of fiscal rules on public finances: Theory and empirical evidence for the Euro area', *CESifo Working Paper*, No.3303.

Musgrave, R. (1959), *The Theory of Public Finance*, Mac Graw-Hill Book Company

Perotti, R. (2013), ‘The Austerity Myth: Gain Without Pain?’, in Alesina, A., Giavazzi, F. (Eds.), *Fiscal Policy after the Financial Crisis* (pp. 307–354), University of Chicago Press.
Reuter, W.H. (2015), ‘National numerical fiscal rules: not complied but still effective?’, *European Journal of Political Economy*, 39, 67–81.
Reuter, W.H., Tkačevs, O. and Vilerts, K. (2018), ‘On the design of stabilizing fiscal rules’, *German Council of Economic Experts Arbeits papier*, No. 11/2018.
Reuter, W.H. (2019), ‘When and why do countries break their national fiscal rules?’, *European Journal of Political Economy*, 57, 125–141.
Sacchi, A. and Salotti, S. (2015), ‘The impact of national fiscal rules on the stabilisation function of fiscal policy’, *European Journal of Political Economy*, 37, 1–20.
Schaechter, A. and Kinda, T., Budina, N. and Weber, A. (2012), ‘Fiscal rules in response to the crisis—Toward the ‘Next-Generation’ rules. A new dataset’, *IMF working paper*, No. 12/187
Sow, M. and Razafimahefa, I.F. (2015), ‘Fiscal decentralization and the efficiency of public service delivery’, *IMF Working paper*, WP/15/59
Tapsoba, R. (2012), ‘Do National Numerical Fiscal Rules really shape fiscal behaviors in developing countries? A treatment effect evaluation’, *Economic Modelling*, 29, 1356–1369.
Thirlwall, A. (2001), ‘The Euro and Regional Divergence in Europe’ inHolmes M. (eds) *The Eurosceptical Reader 2*, Palgrave Macmillan, London.
Warin, T. (2005), ‘The hidden structural features of the fiscal rule: A European saga’, *International Advances in Economic Research*, 11, 29–38.
Wyplosz, C. (2006), ‘European Monetary Union: the dark sides of a major success’, *Economic Policy*, 21, 207–261.
Wyplosz, C. (2013), ‘Fiscal Rules: Theoretical Issues and Historical Experiences’, in Alberto A. and Francesco G. (eds.) *Fiscal Policy after the Financial Crisis* (pp. 495–525), University of Chicago Press, Chicago, 85.

8. Pension spending efficiency

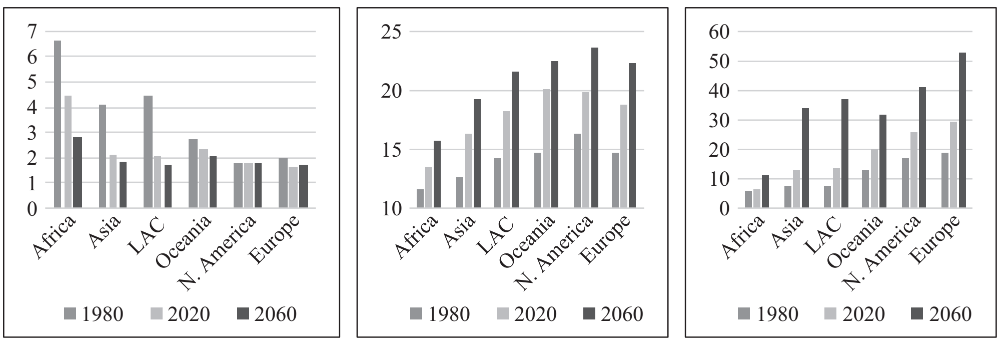
Boele Bonthuis[1]

1. INTRODUCTION

As the world's population ages, striking the right balance between pension adequacy and fiscal sustainability is essential to achieving pension policy objectives while containing cost. As Barr and Diamond (2008) note, '. . . a key concern is to avoid implementing a system that costs more than is necessary to accomplish its objectives, and to balance the level of achievement of those objectives with the costs of achieving them'.

With declining fertility and rising life expectancy the world's population is aging. Since 1980 fertility has dropped by 30–50 percent in Africa, Asia, and Latin America and the Caribbean (Figure 8.1). Fertility rates in Oceania, Northern America and Europe have dropped by less but were already lower in the 1980s. At the same time life expectancy increased by 2 to 5 years across all continents. This has meant that the old-age dependency ratio – the ratio reflecting the number of older people as a share of the working age population – has increased by more than 50 percent in Oceania, Northern America and Europe and even more than 70 percent in Asia and Latin America and the Caribbean. Only Africa saw a minor increase of just 2 percent owing to high fertility rates and modest life expectancy gains.

A. Fertility Rate (number of children) B. Life expectancy at 65 (years) C. Old-age dependency (ratio)



Note: LAC is Latin America and the Caribbean, N. America is Northern America. Old-age dependency ratio is defined as the number of older people (65+) per 100 people of working age (15–64).

Source: United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019, Online Edition. Rev. 1.

Figure 8.1 Fertility rates have fallen, life expectancy has increased leading to increasing old-age dependency ratios

The rise in the old-age dependency ratio is likely to accelerate in the coming forty years. It is expected to increase by more than 50 percent in all continents while more than doubling in Asia and Latin America and the Caribbean. As Figure 8.1 shows, the driving force behind aging going forward is likely to be rising life expectancy rather than falling fertility, with the exception of Africa.

Population aging is one of the driving forces behind rising pension spending, putting pressure on public finances. Pension spending in high-income countries makes up a large part of government expenditure and is the largest social spending item in many countries (OECD, 2020a). On average, advanced economies spend around 7 percent of GDP on pensions. This figure drops to around 3 percent and 1 percent in middle- and low-income countries. In the next decade, pension spending is expected to rise by 0.7 percentage points in advanced economies, 2.0 percentage points in middle-income economies and 0.7 percentage points in low-income developing countries (IMF, 2020).

At the same time the adequacy of pension benefits is still a concern in many countries. Old-age poverty is widespread and coverage of pension systems insufficient in many middle- and low-income countries. And while coverage is generally high in high-income countries, relative old-age poverty is still high in many countries. Moreover, pension reforms of the last decades have led to lower replacement rates (EC, 2018; OECD, 2019b).

To strike the right balance between pension adequacy and fiscal sustainability, pension spending has to be efficient. In other words, for a given level of spending, the value derived by society from the pension system should be maximized while distortions minimized. More formally, we can define a benchmark efficiency frontier, identifying the minimum amount of spending feasible given the outcomes of the pension system. We can define the cost frontier as the minimum spending feasible $c(y_{it}, w_{it}, \beta)$; given outcomes (y_{it}), cost of inputs (w_{it}) and coefficients (β); with a random error (v_{it}), and a measure of cost inefficiency (u_i) representing deviations from the frontier:

$$\ln(\textit{pension spending}_{it}) = \mathrm{c}(y_{it}, w_{it,} \beta) + v_{it} + u_i$$

The cost of inputs reflects both the direct cost of benefits (that is benefit levels) as well as the cost to operate the system.

Efficient pension design depends on social and economic preferences and level of development. Pension policy objectives can be roughly divided into objectives related to insurance against longevity, consumption smoothing and poverty prevention. The degree to which countries want to and can achieve each objective differs. Therefore, a single-most-efficient pension system does not exist. Efficiency can only be measured against varying objectives. Therefore, for the remainder of this chapter we define public pension spending to be efficient if coverage and benefit levels meet the policy objectives of the government without causing undue welfare losses, labor or capital market distortions or tax avoidance.

The remainder of the chapter is structured as follows. First, the objectives of pension systems are explained. Second, pension spending is compared to outcomes of the pension system. Third, pension design and design inefficiency are analyzed and in the last section we briefly conclude. Given the focus on the chapter on public pension spending efficiency, throughout the chapter, pension spending means public pension spending unless stated otherwise.

2. OBJECTIVES OF PENSION SYSTEMS

Pension spending efficiency has to be defined relative to the social objectives of the pension system. Pension policy objectives can be roughly divided into objectives related to insurance against longevity, consumption smoothing and poverty prevention.[2] All of these objectives represent some form of redistribution: redistribution over time and redistribution between people. However, redistribution for its own sake can be seen as an additional objective. Moreover, these objectives are often overlapping, for instance pension credits for unemployment spells redistribute to the unemployed and serve as a form of insurance, consumption smoothing and poverty prevention.

Objectives set for pension policy are by definition normative. There is no 'right' amount of consumption smoothing and there is no 'right' amount of redistribution. However, once objectives are set by policy makers, there are more and less efficient ways to achieve these objectives. The right pension design is therefore important to efficiently achieve these objectives.

Insurance against longevity is an important objective of pension systems. While average life expectancy for an entire cohort can be relatively precisely predicted, on an individual level it is impossible to predict exactly how long someone will live. This uncertainty has a profound impact on retirement planning. If individuals were unable to insure themselves against longevity, they would either run significant risk of running out of funds before eventual death or they would need to keep substantial reserves to be ensured of income in the case of long life. Pension systems typically provide this kind of lifetime uncertainty insurance, whether pension systems are funded, in which case longevity insurance can be provided through annuities, or pay-as-you-go, in which case benefits are typically lifelong by definition.[3]

Consumption smoothing can be characterized as the goal of having a stable path of consumption over someone's life. Considering that most people wish to retire at some point in their lives, this means that income needs to be redistributed over the life cycle. This can either be done through voluntary private savings or through some form of mandatory pension system. Myopic behavior is often cited as an important reason behind mandating participation in a pension system.[4] Diamond (1977) for instance suggests that a sizable fraction of American workers would not save enough in the absence of Social Security. Moreover, an individual's understanding of complex decisions with long time horizons, like retirement planning, is likely to be imperfect (Barr, 2020). It should be noted that it is important to make a distinction between inadequacy of savings because of myopia or the inadequacy of savings due to a lack of resources (Kotlikoff, 1987); two issues that are often conflated because both lead to similar outcomes.

Preventing a lack of resources in old age is another important objective of pension systems. Some people will have low lifetime earnings and are therefore unable to sufficiently save for old age. Given the often lower capacity to work of older people, without adequate savings and without family or community support, they run the risk of falling into old-age poverty, absent of government intervention. Therefore, many countries have introduced elements in their pension systems that provide a minimum living standard for the elderly.

Relieving or preventing old-age poverty leads by definition to redistribution at old age; however, redistribution by itself can be a policy objective. Policy makers might want to reduce inequality at old age, they might want to compensate people for care activities or periods of unemployment during their working lives, provide some insurance against earnings variability or they simply want to acknowledge that certain elements of the cost of life at old age, such as health care and living expenses, are unrelated to lifetime earnings.

3. SPENDING AND OUTCOMES

To assess how efficiently countries achieve their pension objectives, we plot public pension spending against variables reflecting the objectives. The poverty rate and poverty gap are taken to reflect the performance of the pension system in preventing poverty. Benefit levels are taken to assess the ability of the pension system to smooth consumption. And finally, beneficiary ratios, the number of benefit recipients as a share of the population of pensionable age, are taken as a prerequisite for all objectives. If few people receive any pension, poverty prevention, insurance against longevity and redistribution are by definition hard to achieve and consumption smoothing will be solely dependent on the savings behavior of individuals.

It should be noted that these simple correlations do not fully reflect spending efficiency. For instance, all else equal (that is, same average benefit levels, same retirement age, same GDP and so on), a country with an older population will logically spend more on pensions than a country with a younger population. Similarly, all else equal, a country with a higher retirement age will logically spend less on pensions than a country with a lower retirement age, provided that appropriate actuarial adjustments are made for early retirement. Therefore, while these scatterplots reflect where countries stand in terms of spending and outcomes, they are insufficiently detailed to provide a comprehensive assessment of spending efficiency.

3.1 Beneficiary Ratios

Pension beneficiary ratios, the share of people of pensionable age receiving a pension, differ significantly by country. Beneficiary ratios are low in most low-income countries, given the high levels of informality and the lack of social pensions or basic pensions with universal coverage. In Malawi, for instance, only 2.3 percent of the population of pensionable age (60 years) receive a pension (Figure 8.2). These are largely civil servants for whom pensions are more accessible.[5] Conversely, beneficiary ratios in most high-income countries are (close to) 100 percent. However, some high-income countries stand out with beneficiary ratios significantly below 100 percent despite universal coverage of social pensions. For instance, Spain had a beneficiary ratio below 70 percent in 2014. One of the reasons for this is that women often do not receive an individual pension in Spain.[6] The social pension is means tested on a household level, making a person ineligible if the income of a spouse is above the means-testing limit. Moreover, social pensions are disbursed on a household level as well. This means only one person in a household will receive the safety net. Therefore, while pension coverage in Spain is universal, pension receipt is not.[7]

There is a positive correlation between beneficiary ratios and pension spending but significant variability in spending at higher levels of the beneficiary ratio. Countries with low beneficiary ratios typically spend little as a share of GDP. No country with a beneficiary ratio below 20 percent spends more than 5 percent of GDP on pensions (see for instance, Malawi, Nigeria and Peru). Above that range, spending diverges quickly. On the one hand, there are countries spending well below 5 percent of GDP with higher beneficiary ratios (Chile, India, Kenya, South Africa, Thailand and Vietnam for instance). The reasons vary, but for most of these countries a relatively young population helps to keep pension spending as a share of GDP low and in many of these countries public pension benefit levels are relatively low. Finally, in more advanced countries, like Chile, private pensions are widespread with public pensions providing only top ups, keeping public spending relatively low. On the other hand, there are

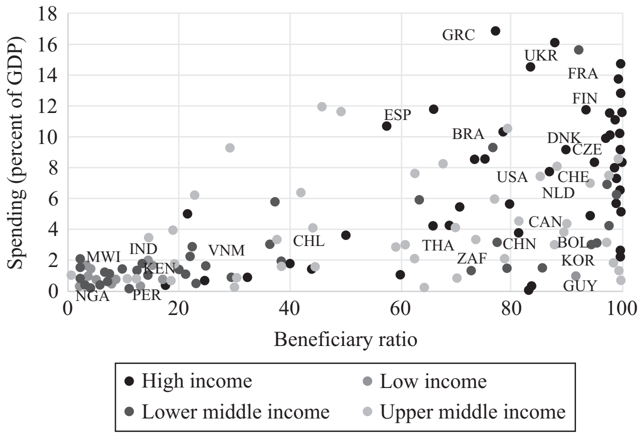


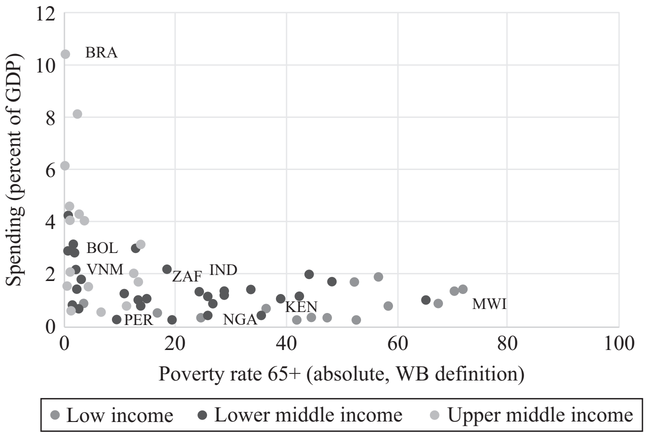
Figure 8.2 Beneficiary ratios

countries with relatively low beneficiary ratios but relatively high spending. Many Balkan counties have relatively high spending because war-related pensions are included in spending but do not show up in beneficiary ratios (that is, the recipients are below the pensionable age). Or there are countries like Spain which have universal coverage and generous benefits but beneficiary ratios below 100 percent. For countries with close to 100 percent beneficiary ratios pension spending differs dramatically. One the one hand, countries like Guyana and Bolivia have residency based basic pensions which cover all pensioners, but benefit levels are relatively low and the population is still relatively young, keeping spending low. On the other hand, there are countries like France, which has a comparatively old population with relatively generous old-age benefits.

3.2 Poverty

Absolute poverty rates for those over 65 years of age, based on a poverty line of less than USD 1.90 per day, are negatively correlated with pension spending. Poverty rates for the population over the age of 65 range from more than 70 percent in Malawi to less than 1 percent in Brazil and Vietnam (Figure 8.3).[8] And while relatively low pension spending can lead to a range of outcomes in terms of old-age poverty, very few countries that spend more have high poverty rates. However, it should be noted that Bolivia, Brazil, Peru and Vietnam have very similar poverty rates but spending ranges from less than 1 percent of GDP to more than 10 percent of GDP. The reason why spending does not perfectly correlate with poverty is that pension spending is not targeted at the poor only, earnings-related pensions and pensions for specific occupations such as civil servants will increase pension spending but are unlikely to significantly influence old-age poverty. Moreover, old-age poverty depends crucially on other factors than just public spending, it depends for instance on household composition, whether people live in multigenerational households or not and on work history, whether the current old were already poor during their working lives or not.

Relative poverty rates might be more relevant for high-income countries and show a stronger correlation with pension spending. As standards of living rise, absolute old-age



Note: The OECD defines relative poverty as disposable income below half the median household disposable income of the total population. The World Bank defines absolute poverty as living on less than $1.90 a day. Observations are for 2015 or latest.

Source: Author's calculations based on data from the OECD, World Bank, International Labour Organization (ILO) and country authorities.

Figure 8.3 Absolute poverty rate

poverty, defined as living on less than USD 1.90 a day, becomes less relevant in reflecting the risk of economic and social exclusion. Therefore, most high-income countries report relative old-age poverty instead, defined as those with income below half the national median equivalized household income. Using this poverty measure there is a stronger correlation with pension spending (Figure 8.4). Countries that spend relatively little report higher relative old-age poverty rates. For instance, China and Korea both have old-age poverty rates of around 40 percent but spend less than 5 percent of GDP on pensions. Both countries have pension systems that have not fully matured yet, albeit to varying degrees. However, even countries with fully matured pension systems like Chile and the United States have relatively high old-age poverty rates (combined with relatively low spending), reflecting in part higher working age poverty and less generous pension systems compared to many European countries. Countries with low levels of relative old-age poverty vary in terms of spending. The Netherlands and Denmark spend less than 10 percent of GDP on public pensions, reflecting the large role private pensions play in both countries, while France spends close to 15 percent.

Poverty rates only reflect one aspect of poverty, apart from the share of people in poverty, it matters how poor these people are. In other words, it matters how far from the poverty line people are removed. This is called the poverty gap. For the World Bank definition of poverty there is no separate data on the poverty gap for people above the age of 65. However, the OECD reports data for those over the age of 65. It shows that, while poverty rates in Denmark and the Netherlands are similar, the poverty gap differs significantly. In Denmark, the elderly poor receive on average income that falls 10 percent short of the poverty line. In the Netherlands, the old-age poverty gap is close to 40 percent (Figure 8.5). For comparison, poverty gaps in Brazil, India and Korea are around 40 percent, albeit for a much larger share of the elderly population.

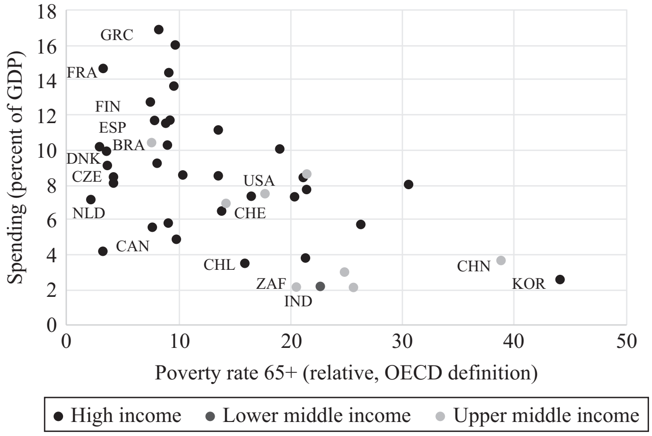
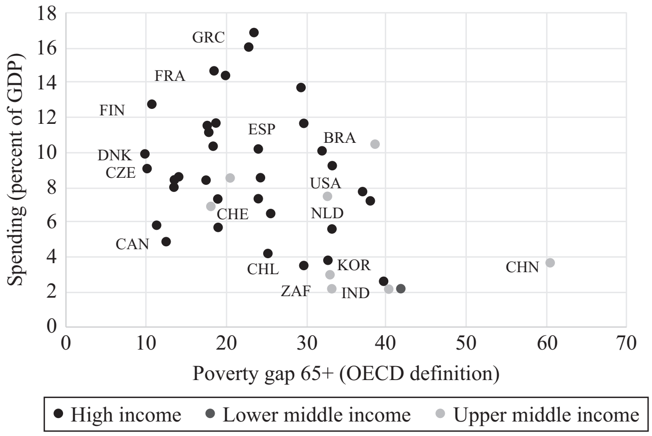


Figure 8.4 Relative poverty rate



Note: The OECD defines relative poverty as disposable income below half the median household disposable income of the total population. The poverty gap is defined as the average distance to the poverty line.

Source: Authors' calculations based on data from the OECD, World Bank, ILO and country authorities.

Figure 8.5 Poverty gap

3.3 Benefit Levels

Not surprisingly, there is a clear positive correlation between benefit levels and spending, but again strong heteroskedasticity. Among countries for which the World Bank has data on social and contributory pension benefits, spending and average benefits (in purchasing power parity terms) are low in India, Kenya and Nigeria for instance (Figure 8.6). On the other hand, spending and average benefits are already a bit higher in Bolivia and China, both countries which expanded their pension coverage significantly in the last decades. Brazil stands out as one of the countries with spending at levels comparable to high-income countries and high benefits

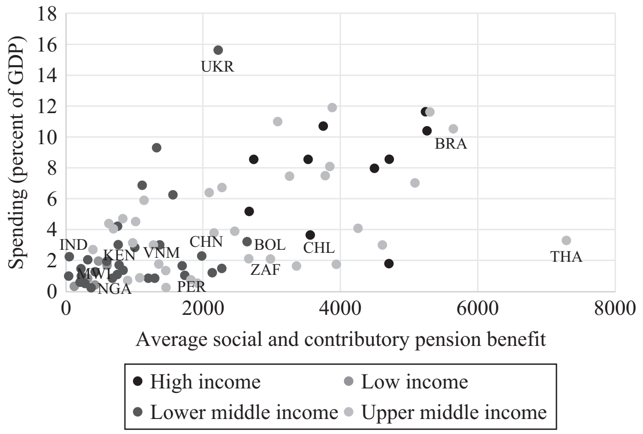
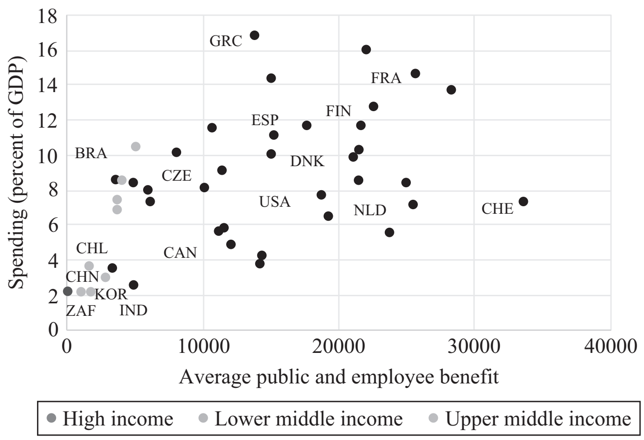


Figure 8.6 Average pension transfer (World Bank definition)

compared to other low- and middle-income countries. However, in Figure 8.7, Brazil mainly stands out as having relatively high spending, while benefits are relatively low compared to high-income countries. For instance, the Czech Republic spends less as a share of GDP but has higher average benefits. Some countries that appear to have high benefits but relatively low spending (Netherlands and Switzerland) have significant private pension coverage, which is not included in pension spending but affects benefit levels. Parsing out private funded benefits from earnings-related benefits is not always possible. For instance, in Finland, earnings-related



Note: Figure 8.6 shows average yearly social and contributory pensions transfers (2005 USD PPP). Figure 8.7 shows current transfers received from public social security and employment-related social insurance schemes (2005 USD PPP). The latter includes private pensions for some countries.

Source: IMF Fiscal Monitor for spending. Authors' calculations based on data from the IMF, OECD and World Bank for benefits.

Figure 8.7 Average pension transfer (OECD definition) pension benefits are partially funded and partially pay-as-you-go, the accrual phase is largely decentralized and managed by company or industry pension funds; at retirement, pension benefits are disbursed by the pension fund of the last employer and the transfer of funds between pension funds is coordinated by a public entity. All of this spending is included in the pension spending measure in this chapter, therefore, excluding earnings-related pension benefits would distort the true picture.

3.4 Efficiency Frontier

To truly assess pension spending efficiency, all outcomes should be tested against spending at once. As described in the introduction, this can be expressed as:

$$\ln(pension\ spending_{it}) = c(y_{it}, w_{it,}\beta) + v_{it} + u_i$$

in which y_{it} would reflect old-age poverty rates, old-age poverty gaps, beneficiary ratios and benefit ratios, w_{it} would reflect cost of running the system, benefit levels but also demographics, which reflect the cost of achieving the stated pension objectives given demographics (that is, it is more costly to reduce old-age poverty in an older population than in a younger one).

In the Appendix the estimation of pension spending efficiency is presented, using the stochastic frontier model for panel data developed by Belotti et al. (2013). It shows that only a very small share of spending is estimated to be inefficient for countries with OECD data. One of the reasons is that this approach quickly runs into perfect collinearity issues. To show this more precisely, pension spending can be rewritten to reflect pension outcomes directly:

$$\frac{pension\ spending}{GDP} = \frac{P_{RA}}{P_{15-RA}} \frac{b}{w} \frac{N}{P_{RA}} \frac{P_{15-RA}}{E} \frac{wE}{GDP}$$

in which $\frac{P_{RA}}{P_{15-RA}}$ is the effective old-age dependency ratio (taking into account the country specific retirement age or effective labor market exit age), $\frac{b}{w}$ is the benefit ratio – relating average benefits to the average wage, $\frac{N}{P_{RA}}$ is the beneficiary ratio – indicating which share of the older population receives a pension, $\frac{P_{15-RA}}{E}$ is the inverse of the employment rate and $\frac{wE}{GDP}$ is the labor share of GDP. This expression can be easily rewritten to reflect old-age poverty of those receiving a pension (Pov_r), the poverty gap (Pov_g), the poverty line (Pov_l) and benefit ratios of those not in poverty (b_{np}/w):

$$\frac{pension\ spending}{GDP} = \frac{P_{65+}}{P_{15-64}} \frac{(1-Pov_g)Pov_l Pov_r + b_{np}(1-Pov_r)}{w} \frac{N}{P_{65+}} \frac{P_{15-64}}{E} \frac{wE}{GDP}$$

The pension spending identity already reflects the different policy targets of a pension system. Policy makers will try to minimize the poverty depth and poverty rate while trying to maximize the coverage ratio and (within reasonable levels) the benefit ratio. However, doing so will directly affect spending. Every dollar not spent on benefits for the non-poor can either be saved (spending will go down proportionally) or be spent on the poor or those not yet covered (spending will stay the same). Since the main policy variables are reflected in the last expression it does not leave anything to estimate. This pension spending identity can decompose the effect of higher spending into the different objectives, but this merely reflects choices between objectives rather than spending efficiency.

In this sense, the assessment of efficiency of pension spending is different than for instance education spending efficiency. In education, spending is channeled through teachers' salaries, school supplies and purchase of services, which leads to certain education outcomes, like literacy rates or PISA scores.[9] In this case spending on teachers' salaries is therefore a means to an end rather than the goal itself. With pension spending, the distinction is less clear. The goal is to provide a pension benefit which is the direct outcome of pension spending. In this case spending is the outcome. This makes frontier analysis of pension spending problematic.

The fact that the spending identity reflects most of the policy targets does not mean there is nothing left to choose as policy makers. There are clear trade-offs between different targets. For instance, countries with low coverage of the general population but high benefits for civil servants could potentially cover more people if civil servant benefits are lowered. Generally, keeping the spending constant in the spending identity, poverty can be reduced and coverage increased by lowering benefits of the non-poor. Raising the retirement age – and therefore lowering the effective old-age dependency ratio – allows for an increase in benefits or an increase in coverage. However, given the trend in demographics, raising the retirement age is likely to merely offset some of the additional required spending to keep benefits and coverage constant.

3.5 Administrative Cost

Administrative cost of a pension system, one aspect of efficiency, is not reflected in the spending identity. One of the earliest international comparisons by Mitchell et al. (1993) shows that (broadly defined) costs vary widely between countries, from 1 percent of benefit expenditure in Norway to 131 percent in Peru in the late 1980s. More recently, Sluchynsky (2015) shows that there is still considerable heterogeneity among countries in terms of administrative cost of pension systems. However, the range has narrowed.[10] Generally, administrative costs exhibit economies of scale (that is, cost per beneficiary falls with rising number of beneficiaries), with benefit management dwarfing contribution collection in terms of cost. Schemes that require management of assets have higher costs on average, most likely due to increased need for skills and more complex organizational structures. Finally, there is a clear link between administrative cost and the level of development in a country, with more developed countries benefiting from lower administrative costs due to better technologies, infrastructure and institutions. However, even among developed countries there is wide variability. In Finland, for instance, administrative cost amounts to 1.7 percent of disbursed pension benefits, most likely because of the relatively complexity of the Finnish pension system with a hybrid funded/pay-as-you-go and private/public management.[11] In the United States, in which pensions are centrally managed the cost of social security is lower at 0.6 percent of total pension benefits.[12] Finally, in the Netherlands, the cost of operating the basic pension amounts to 0.3 percent of disbursed pensions.[13] However, the cost of running the large number of funded defined benefit schemes in the Netherlands is much higher.

Generally, systems with lower administrative costs are considered more efficient. However, there are some trade-offs to be made. Simpler systems (like basic pensions) are cheaper to operate, however, they are more costly on the aggregate because of a lack of targeting. More complicated systems (like hybrid funded/pay-as-you-go and private/public management) are more costly to operate but allow for more targeting, more diversification and potentially more accountability.

4. PENSION SYSTEM DESIGN

Pension systems often contain a combination of distinct components to achieve the above-mentioned objectives. Following the OECD (2019a) pension components can be categorized into three distinct tiers. A first tier provides a minimum level of protection at old age, a mandatory earnings-related second tier aims to replace a share of pre-retirement income and finally a voluntary third tier of private pensions allows for individual retirement savings.

Some types of pension components are naturally more suitable to achieve certain objectives than others within the given constraints. Naturally, first tier pensions are more adept at preventing poverty and providing redistribution, earnings-related pensions provide income smoothing, which is closely linked to consumption smoothing and finally voluntary pensions provide an option to tailor retirement provisions to individual preferences. All of them provide insurance against longevity in different forms with the exception of funded defined contribution pensions taken as lump sums or programmed withdrawals with a short horizon.

The suitability of specific pension components depends on a country's characteristics. Regardless of the stated objectives of pension system design in a country, the size of the informal sector, the administrative capacity of the government, the maturity of the financial system and the level of development more generally are important determinants for the suitability of pension components. There is therefore no one-size-fits-all recipe for achieving pension objectives.

Moreover, pension system design is constrained by other factors as the pension system does not operate in a vacuum. Pension system design can be distortive. For instance, benefits that are not related to contributions can diminish labor supply incentives. On the other hand, pension contributions are part of the total tax wedge and therefore potentially distortive by nature. Because of the long lag between payment of contribution and receipt of benefits, contributions are often perceived as taxes even in pure defined contribution systems.

While some components are more suitable to achieve certain objectives, it should be noted only the entirety of the system will provide the relevant information to assess each objective. For instance, in countries with widespread labor force participation, the earnings-related pension already functions as a significant old-age poverty reduction mechanism. This will have an influence on the appropriate design of the first tier. Conversely, weak formal labor markets influence the viability of any mandatory earnings-related pension; therefore, social pensions or basic pensions in combination with voluntary private pensions for those that can afford it, might, in addition to their objective of old-age poverty prevention, serve as the main (albeit imperfect) consumption smoothing instruments.

4.1 Pension Design and Objectives

4.1.1 Preventing old-age poverty

Many countries have means-tested social pensions providing a minimum level of protection. Someone with no income typically receives the full social pension benefit, while the amount of the benefit is reduced in case of low but positive income or assets (that is, withdrawn against income or assets, for a detailed discussion of means testing see Chapter 14).[14] Social pensions are efficient at targeting the poor at relatively low cost compared to flat rate benefits. Social pensions specifically target poor pensioners since they are means tested. This means that, depending on the accuracy of the targeting, most funds will end up at poor households.

Residency based basic pensions, on the other hand, either pay a fixed amount to everyone who lives or has lived in the country or they pay an amount based on the number of years of residency. If the benefit level is set appropriately, basic pensions can be a powerful tool to prevent old-age poverty. Almost one in five OECD countries have a residency based basic pension. However, basic pensions are also gaining a foothold in developing and emerging economies.[15]

One drawback is that residency based basic pensions are not targeted and therefore cost more at the same level of benefits than a means-tested social pension. For instance, the full benefit level of the means-tested social pension in Belgium is similar to the benefit level of the basic pension in the Netherlands (28 percent and 29 percent of the average wage, respectively; OECD, 2019a). However, in Belgium only 5 percent of the population of pensionable age received a (partial) safety-net benefit compared to more than 100 percent in the Netherlands (with more than 80 percent receiving the full benefit).[16] This meant that, in 2017, Belgium spent a little over 0.1 percent of GDP on the social pension while the Netherlands spent 5.1 percent of GDP on the basic pension. However, even if benefits are flat, they can be partially clawed back through progressive taxation.

Contribution based basic pensions work in a similar way to residency based basic pensions but are only available to those who paid pension contributions. Contribution based basic pensions can therefore only reduce old-age poverty and redistribute among those with a working history. Strict eligibility requirements, like high minimum years of contributions to receive a basic pension (30 years in the Czech Republic for instance), diminish the effectiveness at reducing poverty.

Contributory minimum pensions promise a minimum benefit for those who meet the eligibility criteria for an earnings-related pension but have low entitlements. These pensions function in a similar way as social pensions, that is, they top up benefits to a certain level. However, contrary to social pensions, eligibility depends on a contribution history, which is akin to a contributory basic pension. Like contribution based basic pensions, minimum pensions can only reduce old-age poverty among those with a working history. Unless (formal) labor force participation is near universal, these people are unlikely to be the poorest pensioners. Moreover, since many countries with minimum pensions have relatively strict minimum eligibility requirements, they are also unlikely to be effective at targeting the working poor.[17]

Generally, earnings-related pensions are less adept at preventing old-age poverty than first-tier pensions. But mandatory earnings-related pensions play a role in reducing poverty if, in the absence of these pensions, people would save insufficiently for retirement. Myopic behavior would indeed suggest that this is a risk, which is often cited as an important reason to mandate pension participation. However, earnings-related pensions only reduce poverty for those with a working history. Depending on the generosity of the scheme, pay-as-you-go defined benefit schemes can provide poverty relief for people with low lifetime earnings. However, this would mean that there needs to be significant redistribution within the scheme or that accrual rates need to be high, which is likely to result in fiscal sustainability issues.[18]

4.1.2 Consumption smoothing

While all pensions serving as a minimum level of protection allow for some consumption smoothing, most provide a minimum level of consumption, rather than smoothing per se. Replacement rates from social pensions and residency based basic pensions alone will differ substantially depending on the level of last earnings.[19] Therefore, these pensions alone are unlikely to provide effective consumption smoothing for all income levels in the absence of entitlements to earnings-related pensions.

Benefits in earnings-related pension schemes are based on an individual's earnings history. However, the way earnings translate into pension benefits differs widely among different schemes. Roughly speaking, earnings-related pensions can either be defined benefit, promising a certain pension benefit based on a predetermined pension formula; or defined contributions, in which contributions are predetermined but benefits not. Furthermore, earnings-related pensions can either be funded, benefits are backed by underlying assets, or financed on a pay-as-you-go basis, contributions of the working population pay for benefits of the old. However, in some countries the distinction between these four options is not clear cut and earnings-related pension systems have grown to be of some hybrid form.

An often-heard figure from financial advisers is that retirement income should replace 70 percent of pre-retirement income to provide adequate income smoothing (Biggs and Springstead, 2008).[20] However, income smoothing is not the same as consumption smoothing. Consumption smoothing is likely to require different replacement rates for different earnings levels. The 70 percent income replacement figure is a benchmark for an average earner. Someone who earns less will likely need a higher replacement rate to achieve true consumption smoothing. While retirement reduces some cost of living, like the cost of raising children, commuting, work clothing; other costs will remain constant, like cost of food; and some costs are likely to even increase, like medical expenses. Since low earners typically spend a larger share of their income on necessities, and since health and income correlate, this will mean that low earners will need a larger share of their income replaced than high earners. It also means that optimal replacement rates are dependent on the generosity of other components of the social policy landscape in a country. Myers (1993) reports that replacement rates for the lowest earners in the United States would need to be close to 90 percent while high earners need 60 percent to smooth consumption rather than the 70 percent for average earners. Therefore, to achieve similar consumption smoothing for different income groups, either redistribution or higher private savings for low earners is needed.

Earnings replacement through mandatory pension systems is only one way to reach the necessary income replacement to achieve consumption smoothing. As noted earlier in the chapter, there are good reasons to believe individuals would not save enough to achieve adequate income in old age, making it necessary to mandate participation in the formal pension system. However, this does not necessarily imply that the mandatory pension system should provide the full replacement rate needed to achieve consumption smoothing. Depending on economic preferences in countries, policy makers can choose to leave achieving part of consumption smoothing to the individual. In New Zealand for instance, the basic pension provides a replacement rate of around 80 percent for someone earning half the average wage. Taking Myers' optimal replacement rates for low earners as a benchmark, this would leave around 10 percent replacement to be achieved through voluntary pension arrangements. For average wage earners this would rise to 30 percent.[21]

4.1.3 Longevity insurance

All pension types serve as longevity insurance with the exception of some funded arrangements. Most mandatory pension schemes provide pension benefits for life. However, depending on the choice of payout, funded pension arrangements might not insure against longevity. Lump-sum payments and phased withdrawal leave the individual with the risk of running out of assets before passing away. Phased withdrawal can minimize this risk by choosing sufficiently long horizons, but this would be equivalent to holding large reserves of precautionary savings, compromising the consumption smoothing objective of pensions.

4.2 Prerequisites

Even though different pension components have their innate strengths and weaknesses to deliver on pension objectives, not all types of pensions are suitable in all countries.

The successful implementation of means-tested social pensions depends crucially on accurate identification of the poor, accurate identification of other sources of income and a reliable delivery system for benefits. Social pensions therefore require a certain level of administrative capacity which is often lacking in lower income countries.

The implementation and operation of a residency based basic pension requires less administrative capacity than for a social pension. Residency based basic pensions do not require income information, since eligibility depends only on age, citizenship and/or residency. A citizens' registry and a delivery system are still needed but socio-economic data is not necessary. Administrative capacity needed and administrative cost are therefore likely lower compared to a social pension.[22] However, inadequate registries may make even relatively simple basic pensions difficult to efficiently administer. A common problem, in this regard, is the lack of reliable death registries and the consequent issue of 'ghost beneficiaries'.

Administrative capacity requirements for universal earnings-related pensions are high. While socio-economic data for purely earning-related pensions is not needed. A well-functioning employment history record, a contribution collection system and a benefit disbursement system are needed. This means that widescale use of an earnings-related pension systems is more viable in countries with a large formal sector and sufficient administrative capacity. Moreover, for funded pension systems a well-functioning financial sector is indispensable.

Finally, low administrative capacity can result in low compliance or uptake. Revenue administration and enforcement constraints may also lead to inefficiencies. For instance, the inability to uniformly enforce contribution compliance may lead to higher contributions collected from compliant groups or general revenue financing, the latter implying unintended net taxes or net transfers from/to various income groups driven by the differences across tax and benefit incidences. Lack of information or the means of applying for or accessing benefits may also compromise the efficiency of pension scheme designs.

4.3 Design Inefficiencies

Even when all requirements are met to implement a certain pension system, policy makers should carefully consider unintended consequences of pension design. For instance, the generosity of social pensions can diminish incentives for compliance with contributory scheme rules. The design of these pensions may weaken incentives to fully participate in the pension system or to correctly report contribution-liable earnings. This may lead to both fiscal sustainability and equity issues when benefits are viewed as net transfers and eligibility is difficult to effectively control.

The rate at which the social pension benefit is withdrawn against other income has important implications for the targeting of the system and potential labor supply effects. While social pensions that are fully withdrawn against other income allow for precise targeting, this can lead to reduced incentives to contribute for people with low pension entitlement. After all, for every dollar of additional own pension the social pension benefit is reduced by one dollar. Withdrawing the social pension more slowly against own income can enhance incentives to contribute to earnings-related pensions but reduce targeting.

In the most extreme case, benefits are not reduced against other income at all, as is the case with residency based basic pensions. Since basic pensions are available to all, they are unlikely to significantly affect labor supply incentives.[23] Working longer and accruing more earnings-related pension entitlements or saving more will not influence the level of benefits received from the basic pension. However, similar to the social pension, the generosity of basic pension can diminish incentives for compliance with contributory scheme rules.

A contributory based basic pension or minimum pension that does not lead to extra entitlements for additional years of service also potentially distorts labor incentives. For instance, in the Bahamas, for those with more than 10 years of contributions but with earnings-related entitlements below the minimum pension, there is little incentive to contribute additional years since benefits will not increase.

Final salary schemes or less than full career average earnings are likely to cause unintended redistribution and weaken the link between contributions and benefits. Final salary schemes, which were often a reflection of civil service scheme rules aiming to reward loyalty and long service time, link pension benefits to end-of-career earnings. Final salary schemes can introduce unintended and unwarranted redistribution between people. Those with steep earnings profiles benefit from final salary schemes, while those with flat profiles are hurt by final salary schemes. These schemes can also provide incentives for end-of-career promotions and salary increases unrelated to performance, adding further upward pressure on pension spending.

Inefficiencies in earning-related pay-as-you-go schemes can arise if target benefit levels are unsupported by contribution rates and other pension system parameters. While notional defined contribution systems and most point systems are specifically designed to maintain a strong link between contributions and benefits, in defined benefit systems a divergence between contributions and benefits can easily arise.

In general, the unwarranted special treatment of certain occupations can lead to inefficiencies. Typical cases include the army, civil servants, judges, prosecutors and hazardous occupations (miners, industrial workers, but also, for instance, teachers). These special regimes often are disproportionally generous, especially compared to the total contributions paid. Historically, some of these easements in the case of hazardous work were warranted by working conditions. However, many of these jobs have gradually become easier and safer (or simply ceased to exist) without adjustment to pension rules. Moreover, special regimes for certain occupations are often arbitrary and can lead to unintended sorting into these occupations.

Some groups, like farmers, the self-employed or employees of small enterprises, might be excluded from the mandatory earnings-related pensions. Such exclusions may be driven by either pension policy objectives or by recognized capacity constraints to ensure compliance. Either way, groups excluded from mandatory pension contributions are less likely to save or contribute voluntarily (if provisions are in place to do so), often leading to inadequate pension income. Public schemes operating at a deficit can make these problems worse by requiring partial general revenue financing, using taxes that are also paid by people excluded from mandatory public pension schemes.

Similarly, exempting certain types and brackets of labor income from the contribution base can lead to inefficiencies. Civil servants' pay often includes non-pensionable items, sometimes included to reduce the governments' labor costs and future pension obligations. However, this can weaken the effectiveness of public sector wage setting and can significantly distort the distribution of pension benefits. The use of lower contribution thresholds, often intended to encourage labor supply among new labor market entrants and low wage workers, can undermine coverage expansion and can have unintended redistributive consequences if reflected in inadequate pension benefits in retirement. Conversely, contribution and benefit ceilings can serve to reflect the fact that high earners have an enhanced capacity to save and likely need lower replacement rates for consumption smoothing and are therefore less problematic if set sufficiently high.

Disability pensions are usually the second or third largest pension expenditure item after old-age pensions and sometimes survivor benefits. Inefficiencies can arise from an outdated definition of disability, little focus on reintegration and poor control over disability pension uptake. Inefficiencies may result from both design and implementation issues. Design shortcomings are often related to using only a medical definition of disability without looking at the person's remaining capacity to work and failing to pay sufficient attention to measures promoting reintegration into the labor market. From an implementation perspective, the main issues are poor control over inflow and the difficulty of and limited support for re-entry into the labor market. Well-designed training programs, regular reassessments, anti-discrimination policies and flexible forms of employment can help to reduce benefit uptake and assist people to live an economically independent life.

Finally, expansive eligibility for survivor pensions can lead to inefficiencies. Survivor pensions are usually paid to surviving spouses, direct descendants and sometimes ascendants. While survivor pensions play an important role in providing financial support for those that cannot adjust their labor supply, eligibility is sometimes extended to relatives or to working age survivors irrespective of their long-run labor market prospects. Loose eligibility rules not only generate additional spending pressures but also negatively impact labor supply decisions. Moreover, survivor pension design often leads to unintended redistribution; from singles to couples, from couples with small age differences to couples with large age differences and, if means testing is in place, from dual earner couples to single earner couples (OECD, 2018).

5. CONCLUSION

With declining fertility and rising life expectancy the world's population is aging. The rise in the old-age dependency ratio is likely to accelerate in the coming forty years. Population aging is one of the driving forces of rising pension spending, putting pressure on public finances. At the same time the adequacy of pension benefits is still a concern in many countries.

To strike the right balance between pension adequacy and fiscal sustainability, pension spending has to be efficient. Public pension spending is efficient if coverage and benefit levels meet the policy objectives of the government without causing undue welfare losses, labor or capital market distortions, or tax avoidance.

However, efficient pension design depends on social and economic preferences and levels of development. Pension policy objectives can be roughly divided into objectives related to insurance against longevity, consumption smoothing and poverty prevention. The degrees to which societies want to achieve each objective differ. Moreover, the ability of societies to achieve these objectives differs significantly too.

Pension systems often contain a combination of distinct components to achieve the above-mentioned objectives. A first tier provides a minimum level of protection at old age, a mandatory earnings-related second tier aims to replace a share of pre-retirement income and finally a voluntary third tier of private pension plans allows for individual retirement savings.

Some types of pension components are naturally more suitable to achieve certain objectives than others within the given constraints. Naturally, first tier pensions are more adept at preventing poverty and providing redistribution, earnings-related pensions can easily provide consumption smoothing with defined benefit systems being able to redistribute as well, and finally voluntary pensions provide mainly consumption smoothing. All of them provide insurance against longevity in different forms unless the taking of lump sums is permitted.

The suitability of specific pension components depends on a country's characteristics. Regardless of the stated objectives of pension system design in a country, the size of the informal sector, the administrative capacity of the government, the maturity of the financial system and the level of development more generally are important determinants for the suitability of pension components. There is therefore no one-size-fits-all recipe for achieving pension objectives.

APPENDIX

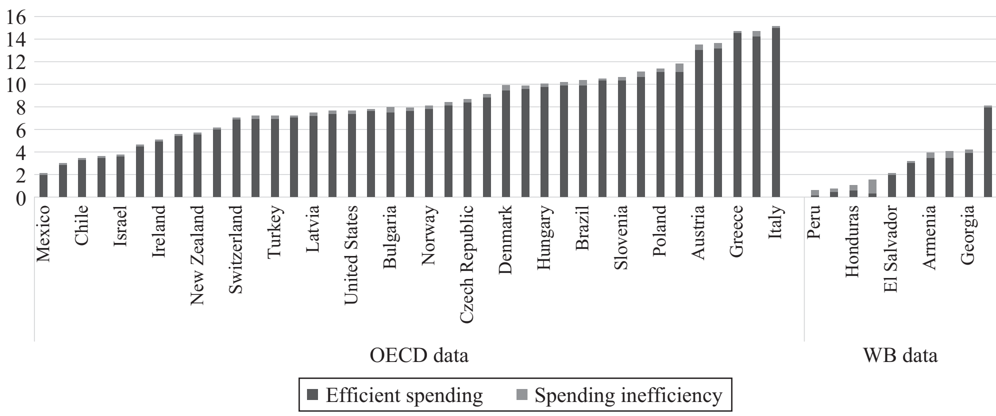
A parametric approach quickly runs into perfect collinearity issues. While all outcomes can be tested against spending at once and corrections can be made for the structure of the population, the more detailed the spending efficiency estimation, the closer to 'estimating' an identity it becomes. For instance, using the stochastic frontier model for panel data developed by Belotti et al. (2013) spending efficiency can be estimated based on the old-age dependency ratio (*OAD*), the poverty rate (*Povr*), the poverty gap (*Povg*), average benefits as share of average income (*b*), not available for all countries, and the average labor market exit age (*LM exit age*):[24]

$$\ln(pension\ spending_{it}) = \alpha_i + \beta_1 \ln(OAD_{it}) + \beta_2 \ln(Povr_{it}) + \beta_3 \ln(Povg_{it}) + \beta_4 \ln(b_{it})\beta_5 \ln(LM\ exit\ age_{it}) + v_{it} + u_{it}$$

v and *u* represent an error term and a strictly positive inefficiency term. Using this equation an estimate of cost efficiency can be made:

$$Efficient\ spending = pension\ spending * \exp(-E(u_{it} \mid v_{it}))$$

Running the above estimation shows that only a very small share of spending is estimated to be inefficient for countries with OECD data (Figure 8.8). For countries with only World Bank data the estimated efficiency is slightly lower but a crucial explanatory variable (average benefits) is missing.



Note: LAC is Latin America and the Caribbean, N. America is Northern America. Old-age dependency ratio is defined as: the number of older people (65+) per 100 people of working age (15–64).

Source: IMF calculations based on data from OECD, World Bank and UN. Average benefits are not included in WB estimation.

Figure 8.8 Spending efficiency

The assessment of efficiency of pension spending through a parametric approach does not seem to have much added value as the chosen variables almost perfectly explain spending. Most countries have close to 100 percent spending efficiency and, for those that do not, there is a clear lack of data, explaining the deficiency. Even for countries with close to 100 percent efficiency the small inefficiency is likely explained by a slight mismatch between the chosen variables and what they are meant to represent. For instance, the old-age dependency ratio is meant to represent the age range of those in retirement versus the age range of those in employment. And while the majority of those in retirement in most countries are likely to be older than 65 it is not a perfect match. Moreover, poverty rates reflect poverty from all income sources, not just public pensions.

NOTES

1. The views expressed in this chapter are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board or IMF management.
2. For a detailed discussion see Barr and Diamond (2008).
3. Benefits in pay-as-you-go systems are technically annuities too, since an annuity is nothing more than the promise to pay a certain amount for life.
4. Mandatory pension participation is not only to ensure consumption smoothing but also to facilitate insurance against longevity, poverty prevention and redistribution. Of which the latter two can only be achieved through some form of government intervention.
5. Plans are in place to expand pension coverage with the introduction of the 2011 pension act to make pension contributions to individual accounts mandatory for private sector workers (in companies with more than five employees) and young/new civil servants. However, this is unlikely to expand coverage significantly given the size of the informal sector and the average firm size in Malawi. According to the ILO, informal employment amounted to more than 80 percent in Malawi in 2013. According to the UNDP around 1 million people were employed in firms with less than five employees in 2012.
6. The gender gap in incidence of receiving an own pension in Spain was 21 percent in 2014 (Lis and Bonthuis, 2019).
7. Moreover, Spain is, in Europe, a popular destination for retirees from other countries, these retirees are unlikely to report receiving a pension from Spain while they might show up in population statistics if they are permanent residents.
8. Poverty rates for the population over 65 years of age are not available for all countries.
9. OECD's Programme for International Student Assessment
10. It should be noted that the definition of cost is not the same between Mitchell et al. (1993) and Sluchynsky (2015).
11. Source: ETK, Finnish Centre for Pensions.
12. Source: Social Security Administration.
13. Source: Sociale Verzekerings bank.
14. The rate at which it is reduced varies by country and sometimes even within a country for different old-age safety nets.
15. Botswana introduced a basic pension in 1996. It provides a monthly pension to every resident citizen over the age of 65 with a valid national identity card. Given the relatively limited qualifying conditions, it is estimated that close to 100 percent of those over 65 receive the basic pension.
16. The basic pension is payable abroad which leads to a beneficiary ratio of more than 100 percent (of the population of pensionable age residing in the Netherlands).
17. Minimum pensions seem to function in many countries as an incentive for longer contribution histories. In these countries, their main objective therefore seems to be to counteract some of the potential negative labor supply effects of mandatory pension contributions.
18. In theory this could be possible in notional defined contribution schemes (high notional interest rate) and point systems (high pension point value) too. However, in reality this is unlikely since notional defined contribution and point systems are typically designed with fiscal sustainability in mind.
19. Among OECD countries an average earner will at best receive a replacement rate of 40 percent in New Zealand from the basic pension (OECD, 2019a).
20. To put this in perspective, the ILO in its Invalidity, Old-Age and Survivors' Benefits Convention stipulates a replacement rate of 45 percent after 30 years of contribution.
21. New Zealand provides fiscally attractive voluntary pension arrangements through its KiwiSaver scheme that can provide the additional consumption smoothing.

22. For instance, Guyana introduced a basic pension in the 1990s. Eligibility requires being age 65 with at least 10 years of citizenship, and at least 20 years of residence. Pensioners can pick up a pension booklet each year containing vouchers for each month of pension payments. With these vouchers pension benefits can be collected at post offices each month. In terms of administrative capacity only a citizens' registry and the necessary logistics to distribute pension booklets are needed.
23. The source of financing is likely to cause distortions though, with contribution financed basic pension likely to be more distortionary than tax financed basic pensions (Auerbach and Kotlikoff, 1987). However, tax financed basic pensions have less of a redistributive effect since the unemployed pay taxes like VAT but generally not contributions.
24. Coverage in many countries does either not have enough observations or lacks variation and is therefore colinear with the constant. Average benefits in the World Bank database does not have enough observations to be included in the estimation.

REFERENCES

Auerbach, A. J., and Kotlikoff, L. J. (1987). *Dynamic Fiscal Policy*, volume 11. Cambridge University Press, Cambridge.

Barr, N., and Diamond, P. (2008). *Reforming Pensions: Principles and Policy Choices*. Oxford University Press.

Barr, N. (2020). *Economics of the Welfare State*. Oxford University Press, USA.

Belotti, F., et al. (2013). Stochastic frontier analysis using Stata. *The Stata Journal* 13 (4): 719–758.

Biggs, A. G., and Springstead, G. R. (2008). Alternate measures of replacement rates for social security benefits and retirement income. *Social Security Bulletin* 68 (2): 1–19.

Diamond, P. A. (1977). A framework for social security analysis. *Journal of Public Economics* 8 (3): 275–298.

European Commission (2018). *Pension Adequacy Report 2018; Current and future income adequacy in old age in the EU*. Vol. Publications Office of the European Union, Luxembourg.

IMF (2020). *Fiscal Monitor: Policies for the Recovery*. Washington.

Kotlikoff, L. J. (1987). Justifying public provision of social security. *Journal of Policy Analysis and Management* 6 (4), 674–689.

Lis, M., and Bonthuis, B. (2019). *Drivers of the Gender Gap in Pensions: Evidence from EU-SILC and the OECD Pension Model*. Social Protection & Jobs Discussion Paper Series, No. 1917. The World Bank, Washington,

Mitchell, O. S., Hsin, P.-L., and Sunden, A. E. (1993). *An Appraisal of Social Security Administration Costs*. World Bank.

Myers, R. J. (1993). *Social Security*. 4th ed. University of Pennsylvania Press, Philadelphia, PA.

OECD (2018). *OECD Pensions Outlook 2018* OECD Publishing, Paris.

OECD (2019a). *Pensions at a Glance 2019: OECD and G20 Indicators*. OECD Publishing, Paris.

OECD (2019b). *Will Future Pensioners Work for Longer and Retire on Less?* Policy Brief on Pensions.

OECD (2020a). *OECD Social Expenditure database*. (www.oecd.org/social/expenditure.htm).

Sluchynsky, O. (2015), Defining, measuring, and benchmarking administrative expenditures of mandatory social security programs. *Social Protection & Labor Discussion Paper Series,* No. 1501.

9. Public−private partnerships, fiscal risks and efficiency[1]

Gerd Schwartz, Özlem Aydin, Rui Monteiro, and Isabel Rial

1. WHAT ARE PPPs?

Public-Private Partnerships (subsequently called PPPs or PPP projects) are projects that are governed by a long-term contract between a government and a private company that covers both investment and service provision. In a PPP, the private company makes an investment in an asset and uses it to provide services to the government or the public, while usually being required to satisfy a set of performance criteria **(see IMF 2021). The legal definitions of what constitutes a PPP may differ from country to country, but the economic performance of PPPs, including their efficiency, can be assessed independent of the legal definition a country uses.**

Infrastructure projects in the form of PPPs have a long history, although details are often scarce. For example, Egypt's Giza Pyramids are said to have been built as PPPs, and the Roman Empire used PPPs extensively, particularly for urban utilities. Some earlier PPPs were based on emphyteutic leases, which are Roman law contracts that are still in use today (for example, in France). Bezançon (2004) describes how PPPs have been used over the centuries in France, including for most bridges in Paris and some other large infrastructure projects (including the *Canal du Midi* and the first Paris Metro lines). Mata (1999, 2002) provides evidence on nineteenth century Portuguese PPPs, including railways, trams, ports, roads and prisons. Similarly, Roberts (2010) describes historical projects in the United States, and de Vries (2013) discusses them in the Netherlands and United Kingdom. PPPs gained popularity in the last decade of the twentieth century, mostly as a result of tight budget constraints and theoretical arguments for increased public procurement efficiency. Some countries, both in Europe and elsewhere, have been particularly heavy users of PPPs (for example, Chile, Colombia, India, Portugal, South Korea and the United Kingdom).

PPPs procure public infrastructure with private financing and bring in private management to provide infrastructure services to the public. PPPs contrast with traditional public procurement, which uses public financing and management. A main argument in favor of PPPs has been that, under certain circumstances, they are more efficient than traditional procurement. There are two basic types of PPP assets (for example, roads, ports, schools, hospitals): *government funded* and *user funded*. With the former, the government provides pre-agreed payments to the private sector to make the asset available (availability payments) or it pays depending on the volume of services provided. With the latter, the assets are funded by user fees, although the government may still subsidize the user fees or the investment, or it may guarantee the private partner's debt or revenue. Combinations of the two funding arrangements are also possible.[2]

2. PPPs AND EFFICIENCY—CONCEPTUAL ISSUES

Infrastructure efficiency refers to creating infrastructure with the highest possible net value for the public. Efficiency in public infrastructure is not simply spending the least to obtain a certain infrastructure volume or extracting the most from a certain amount of public resources. Infrastructure efficiency means that governments should use their scarce resources efficiently (that is, public spending efficiency), spending them on infrastructure assets and services that benefit society the most (that is, economic efficiency). PPPs, as vehicles for procuring public infrastructure, must meet both criteria, that is, public spending efficiency and economic efficiency.[3]

The potential efficiency gains from PPPs are closely associated with having in place good PPP contracts and strong infrastructure governance. PPPs, with their use of private sector capital, knowhow and management to create infrastructure assets and provide infrastructure services, can increase infrastructure efficiency only under certain conditions. More specifically, reaping the potential benefits from PPPs requires:

- **PPP contracts with well-designed risk allocation and management provisions.** PPP contracts are highly complex, long-term agreements (which may run for 25 years or more) between the public sector and one or more private sector counterparts. These contracts cover design, construction or rehabilitation and maintenance of the fixed PPP asset, and, sometimes, its operation by the private partner. They also cover payment agreements, quality standards and performance criteria, and set out responsibilities for managing the asset and bearing contract-related risks throughout the contract's life. Even well-designed and managed PPPs present considerable challenges for governments due to their long-term nature and complex risk sharing and allocation agreements.
- **Strong infrastructure governance to ensure that PPPs are used for the right reasons and that PPP design, implementation and management follow best practice.** PPPs usually require lower initial public-sector spending as they tap into private-sector capital. This can tempt policy makers. Yet, the initial lower fiscal costs can easily come at the expense of much higher future fiscal costs, possible risks to fiscal sustainability and lower overall efficiency. This is the case because PPPs come with additional fiscal risks that can adversely impact future public finances when they materialize. Therefore, selecting the right projects to be executed as PPPs, as well as negotiating and managing PPP contracts, require strong infrastructure governance.

What are the key sources of efficiency gains in PPPs, and what are the challenges for reaping these gains and locking them in? The next sections discuss different aspects of efficiency related to PPPs, focusing on issues related to fiscal risk and fiscal sustainability.

2.1 Efficiency and Contract Design

Well-designed PPPs are based on contracts that encourage a life-cycle approach toward infrastructure assets.[4] In theory, the life-cycle approach helps to lock in efficiency gains. In contrast to traditional procurement, PPPs bind the public and private partners into a long-term contract that bundles infrastructure financing, design, construction, operation and maintenance. Such bundling in the context of a long-term contract can create strong incentives for the private partner to build and manage the asset efficiently, complying with performance criteria during the whole contract period while maximizing investment returns. Since the private partner will operate and maintain the infrastructure asset, it has an incentive to design, build and maintain it to minimize life-cycle costs and maximize efficiency. No such built-in life-cycle incentives exist for traditional public procurement.

Key potential efficiency gains in PPPs relate to better and less expensive maintenance spending. Many countries neglect maintenance spending for infrastructure, given a political bias in favor of new projects that adversely affects the value of existing infrastructure assets (Box 9.1). The PPP life-cycle approach helps eliminate inadequate maintenance, as it is in the interest of the private partner to maintain the asset so as to reduce life-cycle costs while safeguarding performance.

BOX 9.1. THE IMPORTANCE OF MAINTENANCE FOR ACHIEVING INFRASTRUCTURE EFFICIENCY

Poor maintenance adversely affects the value of an infrastructure asset. Yet, since new projects are highly visible, it is often easier to mobilize funds (and both public and political support) for new infrastructure than for protecting the value of an existing asset. In fact, it has been argued that infrastructure maintenance has been so disregarded in some countries that 'build-neglect-rebuild' (BNR, as compared to BOT, 'build-operate-transfer') has become the standard model for infrastructure creation, and at huge cost to the public sector (see PRIF 2013).

PPPs can overcome this challenge. By using a common project-finance methodology for construction and maintenance, a PPP creates both a contractual commitment properly to maintain the asset and the means for meeting this commitment. This includes applying cascading rules for the use of available cash that prioritize maintenance and operational expenditure over other uses (see Delmon 2016 or Yescombe 2002 and 2007 on the role of maintenance reserve accounts in project finance and PPPs).

Yet, reaping the potential efficiency gains requires the quality of infrastructure services (that is, output quality) to be contractible, that is, measurable and feasible to assess in a court of law. Efficiency cannot readily be assessed in all cases. For example, output quality is contractible for highways, and Engel (2020) estimates that the efficiency gains associated with better and cheaper highway maintenance generate government savings somewhere around 10–16 percent of the initial investment. However, output quality is harder to measure and assess in other areas (for example, education), complicating the use of PPPs as vehicles for efficient infrastructure service delivery.

The discipline brought about by using private finance is seen as another potential source of PPP efficiency. In a PPP contract, the private partner's commitment to the project is made credible by putting up funds, raised either from shareholders or markets (for example, loans or bonds), typically known as 'capital at risk'. To do this, the private partner usually has to convince private finance providers of a project's soundness, with future cash flows being credible and adequate to generate a profit margin. This 'due-diligence' process, an integral part of private finance discipline, creates the right incentives and an additional layer of checks that potentially increases PPP efficiency compared to traditional public procurement. For example, in user-funded PPPs, and in the absence of government transfers or guarantees, PPPs can help filter out ill-conceived projects (*white elephants*), given that no private partner will be interested in a project that is not financially sound. However, these potential efficiency gains might not materialize. For example, the risk premia required by the private sector to undertake a PPP may be very high, and governments may be inclined to provide government guarantees to the private partner to reduce these risk premia. Yet, if the government fails to value these guarantees accurately, and prices them into the PPP contract, they may easily reverse the sought PPP efficiency gains when they are called.

PPP efficiency gains can also be achieved through better risk allocation in infrastructure. Risk transfer is one of the main characteristics of PPPs that is critically determined by contract design. Good practice suggests that the project-inherent risks should be borne mainly by the party that is best equipped to control them, while exogenous risks should be shifted to the party that can best absorb or mitigate them. Still, PPP contract design might result in a complex risk allocation structure that is quite different from good practice. For example, poorly designed contract termination clauses might require the government to over-compensate the private partner, with potentially large efficiency losses. Similarly, government guarantees can also impact risk allocation and eliminate the incentives that support PPP efficiency.

Reaping the potential efficiency gains from PPPs also requires addressing different interface risks in PPP contract design. Interface risks arise when infrastructure assets are maintained by the private partner but operated by a public entity (or separate private entity). The interface between the two entities—managing activities related to the same building or asset—may be hard to coordinate. For example, a PPP hospital faces no significant interface risks when activities (for example, building maintenance, clinical services and so on) are managed by the same private partner. However, interface risks are high in PPP hospitals where the private partner is only responsible for maintaining the building and a few other nonclinical services, while public partners (or other private entities) are charged with handling clinical services (Box 9.2).

BOX 9.2. EFFICIENCY AND INTERFACE RISKS IN PPPS

Interface risks arise when not all aspects of project implementation are transferred to a single private partner. It is not uncommon for PPP projects to split project implementation between private and public entities or between different private partners. For example, a given PPP contract may cover the design, construction and maintenance of an asset, but not its operation. This creates potential interface risks, which may be severe, depending on the project in question.

Interface risks vary from sector to sector. For example, if the PPP asset in question is a rail line the interface between design, construction and maintenance on the one hand, and operation on the other, is relatively simple. However, if the asset is a hospital, the interface is usually more complex, particularly when the PPP contract includes some services (for example, cleaning, catering and security) but excludes others (for example, clinical services). In such cases, potential PPP contract violations may be hard to ascertain. For example, is the operating theater contaminated due to poor cleaning or due to faulty surgical practices? As a result, PPP hospitals are often difficult to manage, even in countries with extensive PPP experience.

Sectoral interface risks have been prominent in practice. In France, the Director of the Sud-Francilien hospital center quit, stating he could not run a hospital where the building was overseen by a private entity. Subsequently, the related PPP contract was terminated, and legislation was passed restricting the use of PPPs for hospitals (for example, Law 2014-1653). In contrast, the relative success of PPP hospitals in Portugal may largely be due to the full inclusion of all hospital services in the PPP contract and an effective provision for capping fiscal costs, which mitigate interface risks (see Barros and Monteiro 2016).

Interface risks may also differ from contract to contract and over time. For instance, in both France and the United Kingdom, PPP contracts have evolved over time, encompassing fewer and fewer services. This aligns with gradually acquired evidence (see HMT 2012): when the private provider does not operate end-user services, the inclusion of intermediate services (besides maintenance) in the PPP contract is not efficient.

2.2 Efficiency and Infrastructure Governance

Capturing the efficiency gains from PPPs requires strong infrastructure governance over the full project cycle. Strong infrastructure governance—defined as the institutions and frameworks for planning, allocating and implementing infrastructure investment spending—is essential for ensuring PPP efficiency. A government must have the capacity to identify infrastructure needs and develop strategic plans for addressing them over the long term. It must also allow PPPs to compete for the allocation of limited public resources on a level playing field, to guarantee that projects are procured only when they offer efficiency. Once a project is procured, the government must preserve efficiency through contract management, as well as portfolio risk management and evaluation.

Yet, the long-term nature of PPP contracts provides a major challenge for infrastructure governance. Changes in government priorities or policies, as well as unforeseen demographic, technological or commercial changes, threaten the viability of long-term contracts and may lead to contract renegotiation or collapse. The difficulty of maintaining long-term commitments usually reduces the expected efficiency gains (Box 9.3). This is particularly relevant

BOX 9.3. THE NEED FOR SUSTAINABLE POLICY COMMITMENTS

Many countries have adhered to their long-term PPP commitments even when governments change. In other countries, however, government changes have led PPPs to collapse or resulted in significant contract changes. This has usually resulted in higher fiscal costs. The following examples illustrate the importance of sustainable policy commitments.

In South Africa, political pressures led to abandoning tolls for the Pretoria–Johannesburg highway, requiring the government to compensate the concessionaire for the loss of revenue and the toll equipment already installed.

In France, the government's decision to abandon the plan to collect the *écotaxe* (tolls on heavy goods vehicles) forced an early termination, in 2014, of the PPP contract for toll equipment and collection. The termination required €958 million in compensation to the concessionaire; government's transaction costs for procuring and then canceling the PPP amounted to €70 million (see CC 2017a).

In Scotland, in December 2004, the government abolished the toll for the Skye Bridge, which at one point in time was Scotland's only toll bridge. In the absence of early termination provisions in the contract, the government was obliged to negotiate the bridge's purchase and compensate its operator at considerable additional fiscal costs.

In Portugal, one of the most successful PPP projects, the Vasco da Gama Bridge (completed on time and on budget), underwent a contract renegotiation to fulfill the new government's electoral promise to reduce toll levels. The process was criticized by the Portuguese Supreme Audit Office due to the loss of efficiency that resulted from the negotiation (see TC 2001).

for countries exposed to political instability and/or extreme demand volatility (for example, making user-fee revenues difficult to forecast). Similarly, PPP efficiency is hard to achieve in sectors where government entities operate infrastructure assets exposed to rapid technological changes (for example, IT services). Under these circumstances, the potential PPP efficiency gains are unlikely to be realized in the absence of strong infrastructure governance.

Strong infrastructure requires establishing the laws and institutions to assess all infrastructure projects (including PPPs) on a level playing field. PPP project assessments have to withstand budgetary scrutiny, including for fiscal risk. In this context, strong infrastructure governance requires maintaining a fiscal risk management function in the ministry of finance; establishing a gateway process[5] for project assessment and approval, using international standards in public accounts to ensure fiscal transparency; and integrating PPPs into the annual budget and medium-term fiscal framework.

Special approval processes for PPPs are usually a sign of faulty infrastructure governance. Many countries have adopted separate processes for PPPs that run parallel to the public investment management framework that is applied to traditionally procured projects. This means that PPPs may be assessed and approved under separate methodologies that prevent them from competing with other public investment projects on a level playing field, reducing efficiency in allocating limited public resources. Although many governments have stepped up their scrutiny in recent years, they have not fully integrated PPPs into their overall public investment management frameworks. Yet, this can be done. For example, Cyprus's 2014 Fiscal Responsibility and Budgetary System Law provides for an integrated approach to all public investment projects by subjecting traditionally procured projects and PPPs to the same provisions for preselection, assessment and selection.

Similarly, frequent PPP contract renegotiations are often a sign of weak infrastructure governance. While PPPs are procured expectedly under competition, renegotiation is usually undertaken without competitive pressure and with information asymmetry. Poor negotiating skills often compound low government bargaining power, leading to higher public-sector costs and related efficiency losses (see Guasch 2004 and ECA 2018). Expectations of early renegotiation, originating from poor governance, may also create adverse selection, attracting opportunistic bidders who are not efficient at construction or service provision, but more oriented to game government, and therefore capable of extracting economic rents.

Allowing private-sector entities to submit unsolicited PPP proposals can easily complicate infrastructure governance. Some countries allow private entities, at their own initiative and unsolicited, to submit PPP proposals to the government. If approved, those private entities are entitled to implement the project, or to develop it under some privileged status before a call for tender. Evidence suggests that such proposals usually do not accelerate infrastructure delivery but do create significant fiscal risks and governance issues (World Bank 2018). The often low capacity of public entities to identify, prepare and evaluate infrastructure projects, and incentives to move projects off budget, are often the main reasons for accepting unsolicited proposals. Even when subject to competition, a contract originating from an unsolicited proposal usually does not spur effective competition. Evidence shows that most tenders for unsolicited PPPs attract no competitors, even when the firm originating the proposal has no explicit advantage over potential competitors. This reflects the fact that the originating firm has already done its due diligence, knows the proposed project well and has usually ensured that it recovers its bidding costs whether it wins or loses the tender. Since bidding for a PPP contract is costly, firms naturally refrain from presenting expensive competitive bids when they know that a competitor has a 'first mover' advantage.

Infrastructure governance is also weakened when the PPP partner is not a private entity. In a few countries (for example, Austria, France, Indonesia, China), governments have signed long-term contracts with public corporations to implement public infrastructure projects. These contracts bring a 'business approach' to project management, but often fail to emulate the disciplinary effect of private finance. Private entities are exposed to irreplaceable profit losses and possible bankruptcy when risks materialize. In contrast, public corporations often have multiple objectives and mandates in addition to profitability, making it more difficult to strive for efficiency and avoid moral hazard.

3. THE FISCAL IMPACT OF PPPs

A country's prevailing budgeting and accounting practices are an important factor in understanding and managing the fiscal impact of PPPs. Many countries treat PPPs 'off budget' and 'off balance sheet', that is, they exclude them from the annual government budgets and financial accounting systems. As a result, PPPs do not appear in national public-sector statistics, that is, they are excluded from the fiscal deficit, gross public debt and fiscal risk statements. Such practices tend to reduce efficiency in the provision of infrastructure, introduce biases in procurement decisions (PPP vs. traditional) and affect perceptions of available fiscal space. In fact, poor assessments of PPP-related liabilities and fiscal risks allow inefficient projects to be approved and procured, increasing government liabilities and potentially jeopardizing fiscal sustainability. It is not surprising, then, that a 2018 report by the United Kingdom House of Commons Public Administration Committee (HC 2018) noted that the Treasury had produced no evidence that the PFI program (government-funded PPPs) was worthwhile, 'apart from the fact that it takes debt off the balance sheet'. Similarly, the French Supreme Audit Office has criticized the practice of using PPPs for off-budget purposes (CC 2017b).

3.1 Fiscal Costs and Risks from PPPs

All PPPs create both costs and risks for government. When supporting the private partner, the government may sustain such fiscal costs as explicit commitments or liabilities in the form of viability payments, subsidies, provision of land, acquisition of services and so on (Table 9.1). In addition to fiscal costs, PPPs generate explicit fiscal risks when government liabilities are contingent on the realization of certain events (for example, minimum revenue guarantees)—see Irwin et al. (2018).

Even well-designed PPP contracts create fiscal risks. Since PPP contracts are expected to allocate risks according to each partner's ability to manage the respective risk and cope with its consequences, some risks will necessarily be allocated to the public partner. Therefore, some fiscal risks will arise even from well-designed PPP contracts. At a minimum, a PPP contract must stipulate the financial consequences of an early termination. Typically, this involves buying back core assets and compensating the private partner. In other cases, it involves the PPP company's direct assumption of debt. Contracts may also create fiscal risks through specific clauses detailing the government's responsibility to provide debt and/or minimum

Table 9.1 Fiscal costs and fiscal risks from PPPs

	Fiscal Costs	Fiscal Risks
Explicit	• Liabilities and fixed assets (estimated construction costs) when the government controls the asset and/or bears most of the risks from the contract • Prior or contractually agreed commitments to buy land, resettle people, reestablish utility connections and other infrastructure or compensate affected entities • Up-front capital payments for viability gap funding and other contractually agreed predetermined firm payments during the construction phase • Availability payments, service payments, viability gap funding and other contractually agreed firm payments during the operational phase • Payments by public corporations or subnational governments related to purchase agreements (power, water and so on)	• Explicit guarantees (for example, minimum revenue, exchange rate, reinstatement of economic equilibrium) • Early contract termination (for example, paying for assets, compensating investors, reestablishing service) in case of concessionaire bankruptcy, underperformance, force majeure or public interest reasons (for example, privatization) • Legal disputes • Asset condition at termination
Implicit		• Implicit guarantees: government will strive to maintain infrastructure services

Source: Authors' compilation.

revenue guarantees, exchange rate volatility coverage, reinstatement of economic equilibrium or any other type of support.

Fiscal risks in PPPs can be explicit or implicit. In addition to the explicit risks specified in contracts, PPPs also generate implicit fiscal risks, as the government is the guarantor of last resort for public infrastructure assets and services. Even if not contractually obliged, governments often face political and social pressure to maintain PPP assets in operation. When private partners face financial difficulties, governments may be tempted to rescue them instead of just rescuing the project itself. This practice, or the mere expectation of it, tends to attract opportunistic investors and to foster moral hazard. When implicit fiscal risks materialize, they become fiscal costs. Many highways, railways and ports are currently under public-sector management all over the world because private or PPP projects went bankrupt or could not provide the intended quality of service. Recent examples include the increasing number of governments supporting PPP airport operators during the COVID-19 pandemic, even in cases where operators had not been given an explicit government guarantee.

Implicit fiscal risks are particularly prevalent in PPPs that lack financial viability and/or political sustainability. For example, in user-funded PPPs that lack users, it has not been uncommon for governments to maintain the infrastructure service at considerable fiscal cost (for example, by bailing out private-sector operators that face bankruptcy). Similarly, PPPs that lack political sustainability (for example, when user fees are perceived as unfair) may create unforeseen fiscal costs.

Public corporations can also generate fiscal costs and risks in PPPs, including as off-takers, service providers or finance providers. Some PPPs involve not only the construction of fixed assets (for example, dams, electricity power generation plans), but also require public entities to commit to buying outputs or selling inputs for the project. Particularly relevant are power purchase agreements (PPAs) and similar long-term contracts signed by public corporations for the acquisition of electricity, water, steam, and so on (that is, off-taker agreements). These create firm public commitments to buy goods and services and to pay compensation to the private partner when not buying (for example, due to low demand or cheaper alternatives). Even when PPAs are subscribed by public corporations or subnational governments and not explicitly guaranteed by the central government, they create explicit liabilities for the public sector.

3.2 Budget Flexibility, Fiscal Space and PPPs

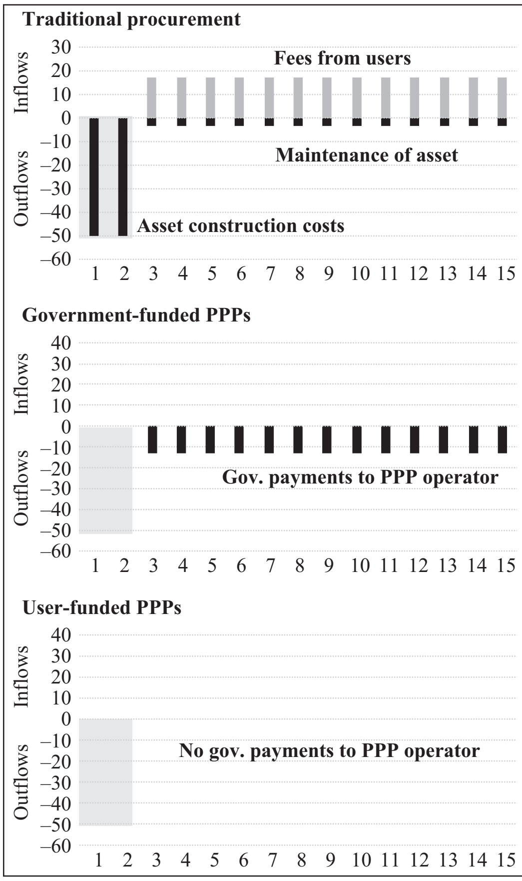
Budgeting for capital projects is always complex. The multiyear, uneven and lumpy nature of capital spending complicates budgeting and accounting for budget spending in a prudent and transparent manner. Ideally, governments should budget and account for capital projects in a way that avoids distorting resource allocation.

PPPs add to the usual challenges of capital budgeting by creating long-term government spending commitments that are, in part, contingent on future events. Compared to traditionally procured public investment, PPPs reduce short-term capital spending at the expense of higher current spending in the future. The upfront private financing of a PPP means that public budgets incur almost no capital spending at the beginning of a project. That is, PPPs shift public-sector costs to the future, either by increasing future government payments (for example, government-funded PPPs) or forgoing future revenue (for example, user-funded PPPs). In contrast, traditional procurement requires particularly high public spending during asset construction. Hence, PPPs change the budget composition, reducing capital expenditure now at the expense of higher current expenditure later. This is illustrated in Box 9.4.

PPPs free up fiscal space in the short term but do not change the government's intertemporal budget constraint. When assets are public, the fiscal costs of procuring them do not disappear and are not absorbed by the private partner, unless unwillingly. Abstracting from efficiency gains, PPPs only change the profile of public cash flows over time but not their net present values. Yet, PPPs allow governments to acquire infrastructure with relatively low initial budget outlays. At the same time, they imply higher future spending, so other future expenditures will have to be reduced (or taxes increased) to pay back the private sector for having financed and created the public infrastructure.

A pervasive use of PPPs may reduce fiscal space and limit fiscal policy responses to external shocks. Initially, PPPs afford governments more room to fund short-term priorities, but at the cost of reducing their room for fiscal maneuver down the road. PPPs often commit future budget resources for the next 25 years or more. As a result, PPPs limit a government's ability to shift spending priorities in the future. In Portugal, for example, the heavy use of PPPs and the resulting ongoing long-term PPP-related spending commitments (which required annual spending of 1.5 percentage points of GDP out of a fiscal deficit of 4.0 percent of GDP in 2011) limited the government's capacity to implement countercyclical fiscal policies in the aftermath of the 2008 financial crisis. Failure to recognize such potential trade-offs may adversely affect fiscal sustainability and PPP efficiency.

BOX 9.4. GOVERNMENT CASH DEFICIT: TRADITIONAL PROCUREMENT VS. PPPS



PPPs create cash flows for government that differ from those created by traditional procurement. To illustrate these differences, we compare the impact on the government's cash deficit using a stylized example of a highway under three options: traditional procurement, government-funded PPPs and user-funded PPPs.

Under traditional procurement the government finances the construction of the highway (in the first two years of the project) through borrowing. For the next 13 years, the government maintains the highway and charges user fees. This generates a large cash deficit during construction. The surplus in the following years depends on the government's ability to impose user fees, and to what extent it maintains the highway.

Under a government-funded PPP, the private partner finances the construction and maintains the highway for the next 13 years. In exchange, the government commits to a stream of payments to the private partner. As a result, no government cash deficit is recorded during construction, while the cash deficits in the following years are larger than they would be under traditional procurement (that is, government payments compensate the private for the highway's construction and maintenance costs, plus a profit margin).

Under a user-funded PPP, the private partner finances, builds, operates and maintains the highway. The government neither makes any payments nor receives any revenues. There are no cash implications for the government.

Note: The example abstracts from efficiency gains/losses in PPPs and makes simple assumptions about asset maintenance costs and fees paid by highway users under each option.

4. PPPs AND EFFICIENCY: THE EVIDENCE

Several data sources offer helpful information on PPPs (for example, financing structures, sectors and regions), but they are often inaccurate and incomplete. A lack of comprehensive data has particularly hampered cross-country comparisons and macro-economic analysis. The academic literature also suffers from lack of data and often relies on a small number of mostly sectoral databases.[6] For example, the World Bank's PPI database—one of the more popular sources of data for analysis—provides some quantitative information on projects but is limited in regional scope and does not provide cost and performance evolution over a project's life cycle.[7] The Appendix to this chapter presents a broader overview on some key sources of PPP-related data.

Available data suggest significant efficiency losses in producing public infrastructure services. The IMF's Public Investment Management Assessment (PIMA)[8] framework, which has been applied to about 65 countries, suggests that losses and waste in public investment are staggering: on average, more than one-third of the resources spent on creating and maintaining public infrastructure are lost due to inefficiency.[9]

However, there is no comprehensive information on the efficiency of PPPs. In most countries, PPP contracts (and contracts related to infrastructure projects in general) are confidential, and only a few countries publish information on project performance. In Europe, most of the disclosed evidence on infrastructure comes from supreme audit institutions and other government audit bodies, while contracts are rarely disclosed. In Latin America, PPP contracts are usually published but performance audits are seldom published; some countries, for example, Chile, publish preliminary studies and periodic performance reports. IMF (2021) presents examples of cases where fiscal risks from PPPs materialized, while recognizing the absence of statistical or systematic quantified information.

Most available data for PPPs focus on effectiveness. Some data suggest that PPPs have much lower cost overruns and delays than traditional public procurement. For example, for Australia, Duffield et al. (2008) found that a sample of 25 PPP projects faced cost overruns of 4.3 percent when measured from contractual commitment to final outcome, while 42 traditionally procured projects had cost overruns of 11.4 percent. They found that PPP projects had significant delays prior to execution, but 'provide[d] far greater cost certainty than traditional contracts'. Portugal's Vasco da Gama Bridge, a €900 million 11 km bridge, was built on time and on budget (as were Portuguese PPP highways in general). Sydney's Cross-City Tunnel also had no cost overruns. However, there are also some counterfactual examples. The Eurotunnel, for example, resulted in cost overruns of roughly 100 percent, mainly due to the channel's difficult geology and rolling-stock issues (see IMF 2021).

There is some country-specific evidence in support of the efficiency of PPPs. Only a few studies, including public spending reviews or reports published by supreme audit institutions, attempt to measure PPP efficiency. In Portugal, the PPP hospital program was evaluated in 2017–19 for public spending efficiency. The PPP hospitals, which had been originally contracted at fiscal costs lower than the Public Sector Comparator (based on historical, not theoretical, costs), provided efficiency gains, delivering better service under a set of health indicators relative to a comparable set of hospitals (see UTAP 2017, 2019a, 2019b).

Yet, several European countries canceled their PPP programs due to inefficiencies deriving from the off-budget treatment of PPPs. In 2012, the United Kingdom treasury recognized that the government funded PFI program had become 'tarnished by its waste, inflexibility and lack of transparency' (HMT 2012) and tried a new PF2 model. In 2017, the United Kingdom's Office for Budget Responsibility considered PFI/PF2 schemes to be a 'source of significant fiscal risk to government'. In 2018, chancellor Hammond confirmed that the PFI/PF2 model had been abandoned. In 2014, France decided to stop hospitals being procured as PPPs. The French Government also canceled its prison and court PPP program in 2018 due to similar efficiency concerns expressed by the French supreme audit office, which claimed that the off-budget treatment, rather than efficiency aspects, was driving PPPs (see IGF 2012, France 2014, CC 2015, and CC 2018).

Similarly, the European Court of Auditors (ECA) found large inefficiencies in a sample of PPPs benefiting from European Union (EU) funding. The ECA report (2018) points to the inadequate analysis and strategic approach towards PPP use. The audit identified weaknesses in the legal and institutional frameworks as the main sources of inefficiency. The report stated that, while 'PPPs have the potential to achieve faster policy implementation and ensure good maintenance standards, the audited projects were not always effectively managed and did not provide adequate value for money. Potential benefits of PPPs were often not achieved, as they suffered delays, cost increases and were under-used, and resulted in €1.5 billion ineffective spending, out of which €0.4 billion were EU funds'.

5. SAFEGUARDING EFFICIENCY IN PPPs—GUIDELINES FOR PRACTITIONERS

Notwithstanding a lack of conclusive evidence on the efficiency of PPPs, there exists practical guidance for achieving efficiency in public investment, including PPPs. Analyses by IMF staff suggest that the average country could halve public investment inefficiencies by adopting best practices in infrastructure governance (see Baum et al. 2020). The high correlation between efficiency and good practice in infrastructure governance suggests that governments can take specific measures to maximize potential PPP efficiency gains.

Specifically, reaping the potential PPP efficiency gains would include the following:

- At the planning stage, ensure that PPPs follow the same public investment management process as those of traditionally procured projects (level playing field).
- At the allocation-of-funds stage, include PPPs in the annual budget and medium-term budget frameworks, properly budgeting and accounting for them to eliminate any bias (integrated processes).
- At the implementation stage, monitor and manage PPP fiscal risks centrally, considering both fiscal costs and risks (comprehensive monitoring).

5.1 Creating a Level Playing Field for PPPs

When planning for infrastructure, PPPs should follow the same public investment management process as those used for traditionally procured projects. This guarantees that all projects are appraised, prioritized and selected using the same social, economic and financial criteria, based on efficiency considerations. Countries should avoid allowing projects to originate as PPPs or following separate PPP pipelines.

Not all infrastructure projects make suitable PPPs. Projects that are procured as PPPs must have certain characteristics that allow for efficiency gains to materialize. First, the project should allow the government to define a set of risks, responsibilities and rights, which can be transferred to a private partner. Second, the project outcomes should be contractible, so that the government can clearly identify performance indicators for the private partner to comply with. Third, project outcomes and overall project sustainability should not be significantly affected by future technological, demographic, or commercial change. Fourth, the project should be based on a politically sustainable policy (for example, user fees) or be robust enough to accommodate likely future policy changes. Indeed, PPPs are suitable for large and complex projects, where private financial and managerial discipline can produce greater efficiency to compensate for higher transactional and financial costs relative to traditional procurement. Roads are an example: their delimitations are well known; standard performance indicators exist (also the long-standing technology for producing roads is not expected to change significantly in the future); and user fees are typically well accepted when new infrastructure brings significant value for users. Hospitals are more complex, but may still be suitable for PPPs, particularly where contractible outcomes can be well defined; mechanisms for dealing with demographic and political change exist; and project delimitation parameters can be established (including clear rules for transfers to/from other hospitals or long-term care facilities, as well as for the entry/discharge of patients).

5.2 Integrating PPPs into Fiscal Processes

At the allocation-of-funds stage, safeguarding efficiency warrants an integrated approach to budgeting, accounting and reporting of PPPs in government accounts. A two-step budgeting process avoids efficiency losses arising from wrong incentives aimed at circumventing budgetary restrictions. Accordingly, an asset's total construction costs should first be appropriated at the time of budget approval, regardless of how the asset is finally procured (for example, as PPP or otherwise). In a second stage, cash appropriations should cover expected payments, if any, to the PPP concessionaire over the contract's term—this will involve extending the horizon of the medium-term budgeting framework, going further than what is usual for capital projects that are procured traditionally. This two-step process ensures that PPPs are subjected to budgetary control and legislative oversight processes that are akin to publicly funded capital projects. For example, in the State of Victoria, Australia, a department considering a PPP must first seek budgetary approval for the capital spending that would be required if the project received public funds. If it is decided to procure the project as a PPP, required budget cash appropriations are planned accordingly, at a second stage, depending on the project cycle (that is, construction or operation) and type of project (that is, government or user funded). In this way, the treatment of PPPs in the budget is similar to that of traditional procurement, allowing for better legislative oversight of long-term PPP commitments. The United Kingdom had already decided to recognize upfront PPP costs in the capital budget at the time decisions are made for projects accounted on budget. As with traditional procurement, the PPPs' costs must be traded off with other capital budget proposals and budgeted under the government unit's fixed budgetary allotment for capital within each separate departmental expenditure limit.

To preserve efficiency over the long term, PPPs should be included in both the annual budget and the government's medium-term budget framework (MTBF).[10] Given that PPP efficiency gains will only materialize over the project's life cycle, their impact should be properly documented over time. Since an MTBF typically only covers three to five years, it cannot provide a full picture of the government's PPP commitments, which can extend as far as 50 years, in some cases. To provide for a longer-term view of public finances, including those related to PPPs, several countries publish complementary fiscal reports, such as the debt sustainability analysis (DSA) or fiscal risks statement (FRS). Typically, these reports provide information on the fiscal impact of PPPs under different scenarios and time frames, depending on the sources of macroeconomic risk and magnitude of the country's PPP portfolio.

Fiscal transparency can help boost PPP efficiency, particularly where national budgeting and accounting systems are not yet aligned with good practices. Disclosing PPP costs and liabilities induces not only greater discipline in public administration, but also provides evidence of waste and inefficiency, allowing governments to introduce corrective measures to preserve efficiency in infrastructure. Most countries establish their budgets on a cash basis, which leads to either excluding PPPs from budget information altogether (for example, user-funded PPPs), or including them only when cash flow takes place late in the project cycle. Some countries produce government financial statements, but they are not necessarily aligned with International Public Sector Accounting Standards (IPSAS), which prescribe the inclusion of most PPPs in the government accounts.[11] Thus, in many countries, PPPs remain broadly unreported, making it difficult to assess efficiency in a systematic manner. Some countries find alternative ways to improve PPP transparency. For instance, Chile publishes an annual contingent liability report, which originally presented information on contingent liabilities from revenue and exchange rate guarantees provided to PPP operators. Since 2016, this report has been expanded to include other types of contingent liabilities and better probability estimates of contingency materialization. Portugal's ministry of finance publishes quarterly data on PPPs, including the time profile of expected future payments for the existing PPP portfolio. In the Philippines, the ministry of finance publishes an annual Fiscal Risk Statement that provides a comprehensive view of the country's exposure to macroeconomic risks and contingent PPP-related liabilities (among other sources of fiscal risk) and summarizes the government's management and mitigation policies.

Fiscal transparency in PPPs can be increased gradually over time, fostering efficiency in infrastructure. Improving national budgetary and accounting practices and developing the technical skills to assess and monitor complex PPP projects might take time. In the near term, governments can rely on analytical tools that focus on big picture estimates. The PPP Fiscal Risk Assessment Model (PFRAM 2.0), developed by IMF and World Bank staff, can be used to assess the potential fiscal costs and risks associated with PPPs. PFRAM 2.0 helps governments identify and quantify the macrofiscal implications of PPPs, understand main sources of fiscal risk and discuss potential mitigation measures (Box 9.5).

BOX 9.5. PPP FISCAL RISK ASSESSMENT MODEL (PFRAM 2.0)

Since it was first launched in 2016, the PFRAM has been used in the context of IMF and World Bank capacity development activities, and by country authorities, to assess PPP projects' long-term fiscal implications. In practice, assessing PPPs involves gathering specific project information and making judgments about the government's role at key stages of the project cycle. PFRAM 2.0 provides a structured process for gathering information for a PPP project portfolio on a simple, user-friendly, Excel-based platform, following a five-step decision tree:

- **Who initiates the project?** The impact on the main fiscal aggregates (that is, deficit and debt) vary depending on the public entity ultimately responsible for the project (for example, central/local government, public corporation and so on).
- **Who controls the asset?** Simple standardized questions assist the user in making an informed decision about the government's ability to control the PPP-related asset through ownership, beneficial entitlement, or otherwise.
- **Who ultimately pays for the asset?** The project's funding structure determines its implications for the main fiscal aggregates. PFRAM 2.0 allows for three funding alternatives: (a) the government using public funds; (b) the government allowing the private partner to collect fees directly from users (for example, tolls); or (c) a combination of the two.

- **Does the government provide additional support to the private partner?** Governments can directly fund PPPs but can also support the private partner in various ways, including by providing guarantees, equity injections, or tax amnesties.
- **What does the PPP contract risk allocation tell us about macrofiscal risk?** Understanding the macrofiscal implications of a PPP contract and the potential need for government action can, in practice, be challenging. PFRAM helps analysts identify main fiscal risks arising from a PPP contract and provides a framework for evaluating risk likelihoods and fiscal impacts. It also enables analysts to assess the need for risk mitigation measures and/or government actions.

PFRAM 2.0 generates standardized outcomes for either a portfolio of PPP projects or a single one, based on project-specific and macroeconomic data. These include (i) projected cash flows; (ii) cash- and accrual-based fiscal tables/charts; (iii) debt sustainability analysis including and excluding the PPP projects; (iv) analysis of the main fiscal aggregates' sensitivity to changes in macroeconomic and project-specific parameters (for example, contract termination); and (v) a summary risk matrix of each project.

Source: The IMF's 'Infrastructure Governance' webpage includes links to PFRAM Excel tools in English (https://www.imf.org/external/np/fad/publicinvestment/data/pfram2english.xlsm) and French (https://www.imf.org/external/np/fad/publicinvestment/data/pfram2french.xlsm), as well as the user manual (https://www.imf.org/external/np/fad/publicinvestment/pdf/PFRAM2.pdf).

5.3 Centrally Monitoring and Managing PPPs

Using an integrated 'whole government' approach, governments should monitor and manage individual PPPs and the overall PPP portfolio over the entire projects' life cycles. One of the government's key objectives should be understanding and managing PPPs' fiscal costs and risks in order to preserve efficiency gains. This is a complex undertaking, as different agencies (for example, line ministries, subnational governments, public corporations) are often responsible for managing individual PPP contracts during project identification, construction and operation. For example, contracting agencies are often not mandated to look at fiscal risk arising from the overall PPP portfolio (that is, the consolidation of all existing PPP projects either implemented or in the pipeline). This indicates a need for information sharing across agencies.

PPP fiscal risks should be assessed (that is, quantified and accepted or rejected) at project inception, and then monitored and managed closely during the whole project life cycle. This should be the responsibility of the ministry of finance (or budgetary authority) as it is ultimately responsible for safeguarding long-term fiscal sustainability. The ministry of finance should be able to identify, estimate and manage fiscal costs and risks arising from PPPs to ensure that efficiency gains are retained during the project's life cycle. To this end, the ministry of finance should have a clear mandate prescribed by the legal and regulatory framework, oversee a strong gateway process to stop any project deemed as unsustainable or too risky and possess the technical skills to identify, estimate and mitigate fiscal risks. This function can be implemented in different ways within the ministry. Some countries have created specialized units (for example, PPP units, investment units, fiscal risk units) with the necessary technical skills to manage fiscal risks associated with infrastructure projects, including PPPs. Other countries have charged existing ministry of finance departments or units (for example, budget department, debt management office and so on) with this task.

It is important for risk management to cover the overall PPP portfolio. Fiscal risks from PPPs tend to correlate among themselves and with other public investment projects (for example, network effects), and are usually highly dependent on key macroeconomic variables (for example, GDP growth, the nominal exchange rate). Therefore, ministries of finance should assess and take decisions at the portfolio level to account for project correlation and evaluate projects as part of a system rather than in isolation (see Monteiro et al. 2020). This is key to boosting efficiency both at the project and network levels and to ensuring that efficiency is preserved over time.

5.4 Safeguarding Efficiency in PPPs—Concluding Thoughts

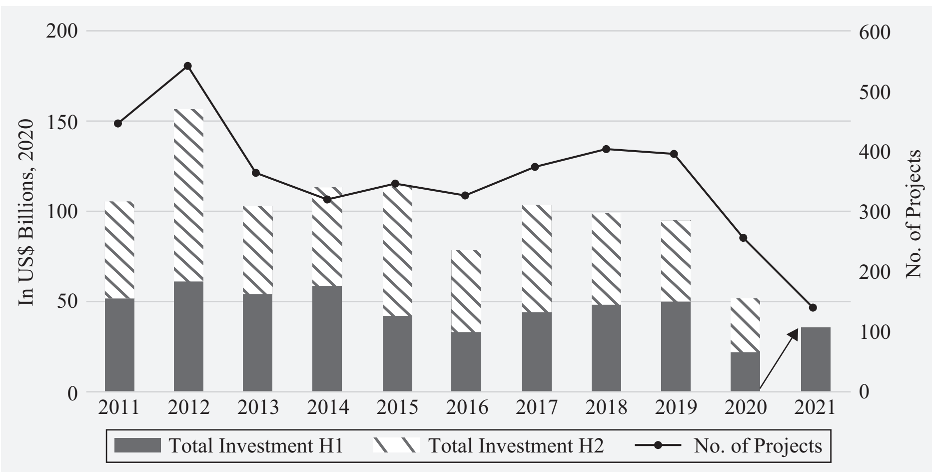
PPPs are not a panacea for increasing the efficiency of public infrastructure services. Governments can take measures to safeguard efficiency in PPPs by focusing on best practice in PPP contracting and strengthening the infrastructure governance frameworks that support them. The latter includes measures to ensure that PPPs are fully integrated in the budget planning and allocation processes, and that the fiscal costs and risks arising from PPPs are centrally monitored and managed. Without adequate fiscal risk management in the context of a strong governance framework, PPPs will, in practice, tend to be inefficient, even when deemed efficient ex-ante. A strong fiscal risk management function in the ministry of finance can support the selection of projects and contracts, aiming at reaching efficiency and not at bypassing fiscal controls, and at protecting that same efficiency over the life of the PPP projects.

APPENDIX: DATA SOURCES ON PPPs—WHAT IS AVAILABLE AND WHERE TO FIND IT

Understanding strengths, shortcomings and pitfalls of the available data is important for all serious analytical work on PPPs. This appendix presents an overview of selected PPP data sources.

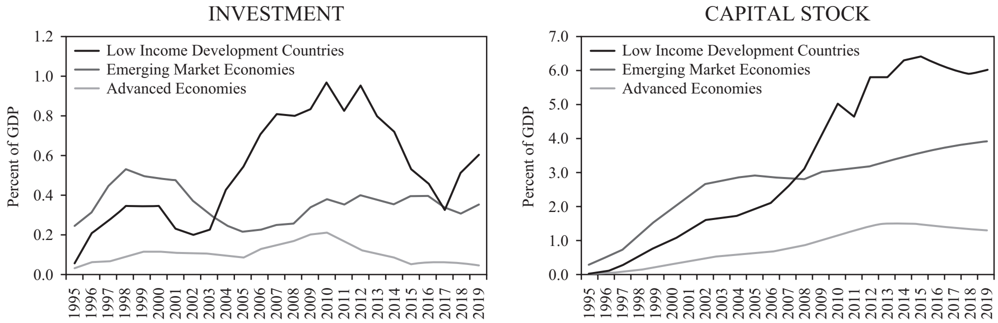
The World Bank's Private Participation in Infrastructure (PPI) database has a broad sectoral coverage, although it is far from being comprehensive or accurate.[12] The database includes projects in energy, transport, water and sanitation, municipal solid waste, natural gas transmission and distribution and information and communication technology (for example, fiber optic cables, mobile towers). Data are available since 1990 and cover 6400 infrastructure projects in 137 countries. Accordingly, investment commitments for the first half of 2021 account for US$35.6 billion across 133 projects (Figure 9.A1). Notwithstanding its wealth of useful project data, the database lacks comprehensiveness and is ill suited for broader analysis. For one, it covers only PPPs in low-income developing markets (LIDCs) and developing economies (EMEs). Also, it excludes PPPs in social sectors (for example, schools, hospitals, prisons, housing). Similarly, quality and internal consistency of the data vary. For example, only projects that are designated as PPPs are included, and projects are being left out where the national definition of PPPs is not aligned with good practices. Also, renegotiations of projects are not treated consistently across the dataset, which results in some double counting of projects and a loss of information on originally agreed costs and deadlines.

The IMF's Investment and Capital Stock (ICS) Dataset[13] includes public, private and PPP-related investment flows, and estimates of real public capital stocks for 1990–2019 for about 170 countries (Figure 9.A2). The data show a particularly rapid increase in PPP investments in low-income developing countries (LIDCs) until 2010, when they peaked at over



Sources: PPI_2021_Half-Year-Report.pdf (worldbank.org).

Figure 9.A1 Investment commitments in infrastructure projects with private participation in low-income developing countries and developing economies, 2011-H1 2021



Sources: https://infrastructuregovern.imf.org/content/dam/PIMA/Knowledge-Hub/dataset/WhatsNewinIMFInvestmentandCapitalStockDatabase_May2021.pdf' https://infrastructuregovern.imf.org/content/dam/PIMA/Knowledge-Hub/dataset/WhatsNewinIMFInvestmentandCapitalStockDatabase_May2021.pdf

Figure 9.A2 PPP investment and capital stock, in percent of GDP

1 percent of GDP in 2010, and a sharp contraction during 2013–17. PPP investments remained more stagnant in EMEs and never really took off in advanced economies (AEs). By 2019, PPP capital stocks averaged 7 percentage points of GDP in LIDCs, 4 in with EMEs and about 1.5 in AEs. While the ICS dataset includes a broader set of countries (including AEs), it also excludes PPPs from social sectors. Similarly, it also suffers from the PPI's shortcomings in terms of quality and internal consistency, also as it uses PPI data for estimating public investment and capital stock in LIDCs.

Other databases often have a more limited regional coverage. Notably, a ***dataset available from the European Investment Bank's PPP Expertise Centre (EPEC)*** offers information on the total value of 1799 PPP projects by economic sector for 1990–2016, but is limited to the EU-27, the UK, Turkey and Western Balkan countries.[14] Yet, it also includes projects in the social sectors (education, recreation and culture, healthcare, and housing and community services), and offers fairly consistent data of good quality.

Several other publicly available resources provide mostly qualitative information on PPPs. For example, the *Economist Intelligence Unit's (EIU) Infrascope* offers a standardized index to access the ability of various countries to carry out PPP transactions in a sustainable manner. So far, the index has been applied in Africa, Asia-Pacific, Eastern Europe and Latin America and the Caribbean.[15] Similarly, the World Bank publishes information on PPP regulatory frameworks in its annual reports on *Procuring Infrastructure PPPs*. It benchmarks regulatory frameworks around the world against internationally recognized good practices in procuring PPPs, and identifies areas for improvement in the preparation, procurement and management of PPPs.[16]

In addition, several databases on PPPs that mostly offer detailed project-by-project data are only available to subscribers. This includes data collected by Prof. Bent Flyvbjerg at Oxford University and IJ Global. The latter includes detailed global data on PPP-related transactions and assets covering key industries, including oil and gas, renewables, power, transport and defense.[17]

There are also several databases that report on infrastructure projects more broadly (not just PPPs), all of which suffer from significant shortcomings. For example, the *Global*

Infrastructure Project Pipeline (*'Pipeline'*) contains information on public infrastructure projects in 6 sectors (that is, energy, communications, transport, water and waste, social and 'others') and is built around 8 standardized project stages through the project life cycle. The *Pipeline* has a promotional investor-oriented focus, aiming to capture relevant early-stage project information. At end-2021, it contained data on infrastructure projects in 56 countries across 5 regions with a total value of US$224 billion; all data are sourced directly from government authorities.[18] The data suffer from quality issues that come with self-reporting (for example, a lack of quality control); incomplete coverage (for example, limited variables and country coverage);[19] and infrequent updating (for example, only if and when infrastructure projects move from one defined project stage to the next). Not unlike the *Pipeline*, *SOURCE*, an infrastructure project development and data collection platform for both traditionally procured and PPP projects, seeks to provide project data across the investment cycle, covering governance, technical, economic, legal, financial, environmental and social issues. At end-2021, SOURCE's stated coverage included 621 projects from 79 countries, although data on only 41 projects were publicly accessible; of these, 25 were still in the project definition phase and 21 came from a single country (Uzbekistan).[20] The available data lack comprehensiveness and reliability, that is, what is available depends on what project users decide to enter and in what form, and if users decide to update the data when relevant information changes.

NOTES

1. Prepared by Gerd Schwartz, Özlem Aydin, Rui Monteiro and Isabel Rial (all International Monetary Fund, IMF). The views and opinions expressed here are those of the authors and do not necessarily represent the views of the IMF, its Executive Board or IMF management.
2. Some countries distinguish between PPPs (government-funded projects) and concessions (user-funded projects). We define both as PPPs, given their common contract arrangements, which combine private sector financing and management.
3. Efficiency in public procurement is sometimes presented as *value for money (VFM)*, a concept that encompasses the effective, efficient and economic use of resources (see World Bank 2016). As effectiveness is a necessary condition for efficiency, in theory, *VFM* encompasses both *economic efficiency* and *spending efficiency*. In practice, however, the term *VFM* has been closely associated with the methodology for assessing VFM, which was developed in the United Kingdom and has recently been cited as a source of inefficiency in infrastructure procurement (NAO 2018). To avoid confusion, we will not use the term *VFM* here, but instead refer to the different aspects of efficiency.
4. Singh (2018). A full life cycle for a project encompasses the following phases: identification, preparation, appraisal, detailed design, procurement, implementation (which includes construction and operation), contract management, monitoring and ex-post evaluation.
5. A gateway process for a project is a formal sequence of approvals, or 'gates', where the Finance Ministry can stop a project that is deemed unaffordable or creates too much fiscal risk. Approval gates usually include a preliminary project approval, an approval of the draft contract and tender rules before the call for tender, an approval of the contract before contract close and an approval of any change to the contract.
6. See, for example, Duffield et al. (2008); Flyvbjerg (2014); and Saussier (2013).
7. See https://ppi.worldbank.org/en/ppi for public access to the World Bank's PPI database.
8. The PIMA is a comprehensive framework to assess infrastructure governance for countries at all levels of economic development. PIMAs evaluate the procedures, tools, decision making and monitoring processes used by governments to provide infrastructure services to the public. PIMAs also help identify reform priorities related to infrastructure governance and devise practical implementation steps. See https://infrastructuregovern.imf.org/.
9. IMF (2015), and Baum et al., chapter 3 in Schwartz et al. (2020).
10. To access fiscal policy in the medium term, most governments have developed medium-term budget frameworks (MTBFs). An MTBF clarifies the level of future government commitments, including those related to capital investment projects and PPPs, to ministers, the legislative and the public in general.
11. Under IPSAS, when certain conditions are met, the impact of PPPs on fiscal deficit and debt measures is like those of traditionally procured projects, reducing the bias in favor of PPPs. See IPSAS 32 - *Service Concession Arrangements: Grantor* at https://www.ifac.org/system/files/publications/files/B8%20IPSAS_32.pdf.

12. See https://ppi.worldbank.org/en/ppi
13. See https://infrastructuregovern.imf.org/content/dam/PIMA/Knowledge-Hub/dataset/WhatsNewinIMFInvestmentandCapitalStockDatabase_May2021.pdf
14. See https://data.eib.org/epec
15. EIU Infrascope reports are available at https://ppiaf.org/documents/2399/download (2015); https://library.pppknowledgelab.org/documents/2400/download (2014); https://www.ebrd.com/downloads/news/eecis.pdf (2012); https://library.pppknowledgelab.org/documents/2403/download (2014)
16. See https://library.pppknowledgelab.org/documents/5453/download/ (2018)
17. See https://ijglobal.com/data/search-transactions
18. https://pipeline.gihub.org/
19. For instance, at end-2021, the Pipeline reported on 86 projects from Mexico, while the Mexican government's own investment and infrastructure projects hub had data on a total of 521 projects (185 new projects and 336 projects in operation). https://www.proyectosmexico.gob.mx/en/home
20. https://public.sif-source.org/source

REFERENCES

Barros, Pedro Pita & Rui Sousa Monteiro (2016) 'Public-Private Partnerships', in *World Scientific Handbook of Global Health Economics and Public Policy* (edited by Richard M. Scheffler), pp. 351–404.

Baum, Anja, Tewodaj Mogues & Geneviève Verdier (2020) 'Getting the most from public investment', Chapter 3 in Schwartz et al. (editors, 2020) *Well Spent: How Strong Infrastructure Governance Can End Waste in Public Investment*, Washington, DC: International Monetary Fund. (https://www.elibrary.imf.org/view/IMF071/28328-9781513511818/28328-9781513511818/ch03.xml)

Bezançon, Xavier (2004) '*2000 ans d'histoire du partenariat public-privé pour la réalisation des équipements et services collectifs*', Paris: Presses Ponts et Chaussées.

CC 2015 – France, Cour des Comptes (2015) 'Les partenariats public-privé des collectivités territoriales: des risques à maîtriser', in *Rapport public annuel 2015 Tome I*, Paris, February. (https://www.ccomptes.fr/sites/default/files/EzPublish/125-RPA2015-partenariats-public-prive.pdf)

CC 2017a – France, Cour des Comptes (2017) 'Ecotaxe poids lourds', in *Rapport public annuel 2017 Tome I*, Paris, February. (https://www.ccomptes.fr/sites/default/files/EzPublish/05-ecotaxe-poids-lourds-Tome-1.pdf)

CC 2017b – France, Cour des Comptes (2017) *La politique immobilière du ministère de la justice: mettre fin à la fuite en avant*, December. (https://www.ccomptes.fr/sites/default/files/2017-12/20171213-rapport-immobilier-ministere-justice_0.pdf)

CC 2018 – France, Cour des Comptes (2018) 'Dix ans après le lancement de l'opération Campus, un premier bilan en demi-teinte', in *Rapport public annuel 2017 Tome I*, Paris, February. (https://www.ccomptes.fr/sites/default/files/2018-01/11-dix-ans-apres-lancement-operation-Campus-premier-bilan-demi-teinte-Tome-1.pdf)

De Vries, Pieter (2013) 'The Modern Public-private Demarcation: History and Trends in PPP', in *The Routledge Companion to Public-Private Partnerships*, (edited by Pieter de Vries and Etienne B. Yehoue), pp. 9–28, Abingdon: Routledge.

Delmon, Jeffrey (2016) *Private Sector Investment in Infrastructure: Project Finance, PPP Projects and PPP Frameworks*, Alphen aan den Rijn, Netherlands: Wolters Kluwer.

Duffield, Colin, Peter Raisbeck & Ming Xu (2008) *Report on the performance of PPP projects in Australia when compared with a representative sample of traditionally procured infrastructure projects*, Research Report, University of Melbourne. (https://www.infrastructureaustralia.gov.au/sites/default/files/2019-06/PC_Submission_Attachment_K.pdf)

ECA 2018 – European Court of Audits (2018) *Public Private Partnerships in the EU: Widespread shortcomings and limited benefits*. (https://www.eca.europa.eu/Lists/ECADocuments/SR18_09/SR_PPP_EN.pdf)

Engel, Eduardo, Ronald Fisher & Alexander Galetovic (2020) *When and How to Use Public-Private-Partnerships in Infrastructure: Lessons from the International Experience*, NBER. (https://www.nber.org/papers/w26766)

Flyvbjerg, Bent (2014) 'What You Should Know about Megaprojects and Why: An Overview,' *Project Management Journal*, vol.45, no.2, April-May, pp. 6–19. (https://arxiv.org/ftp/arxiv/papers/1409/1409.0003.pdf)

France (2014) *Loi n° 2014-1653 du 29 décembre 2014 de programmation des finances publiques*. (https://www.legifrance.gouv.fr/loda/id/JORFTEXT000029988689/2020-10-29/ and https://www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000036535156/2018-01-01/)

Guasch, José Luis (2004) *Granting and Renegotiating Infrastructure Concessions: Doing it Right, WBI Development Studies,* Washington, DC: World Bank. (https://openknowledge.worldbank.org/handle/10986/15024)

HC 2018 – House of Commons Public Administration and Constitutional Affairs Committee (2018) *After Carillion: Public sector outsourcing and contracting*, July, Seventh Report of Session 2017–2019. Report, together with formal minutes relating to the report. (https://publications.parliament.uk/pa/cm201719/cmselect/cmpubadm/748/748.pdf)

HMT 2012 – United Kingdom Treasury (2012) *A new approach to public private partnerships*, December. (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/205112/pf2_infrastructure_new_approach_to_public_private_parnerships_051212.pdf)

IGF 2012 – France, Inspection Générale des Finances (2012) *Évaluation des partenariats public-privé* (rapport par Jean-François Juéry, Emmanuel Monnet, Justine Coutard, Adrienne Brotons et Alban Hautier), December. (http://www.igf.finances.gouv.fr/files/live/sites/igf/files/contributed/IGF%20internet/2.RapportsPublics/2012/2012-M-027.pdf)

IMF 2015 – International Monetary Fund (2015) *Making Public Investment More Efficient*, Washington, DC.

IMF 2021 – International Monetary Fund (2021) *Mastering the Risky Business of Public-Private Partnerships in Infrastructure,* Washington, DC.

Irwin, Timothy C., Samah Mazraani & Sandeep Saxena (2018) *How to control the fiscal costs of public-private partnerships*, How-To-Note 18/04, Washington, DC: International Monetary Fund. (https://www.imf.org/~/media/Files/Publications/HowToNotes/howtonote1804.ashx)

Mata, Maria Eugénia (1999) 'A Companhia das Obras Públicas de Portugal', *Estudos de Economia*, Vol.XIX, no.1 (Inverno). (https://core.ac.uk/download/pdf/303770861.pdf)

Mata, Maria Eugénia (2002) 'Institutions and business Henry Burnay: A case study', mimeo, XXII, Encontro Anual da Associação Portuguesa de História Económica e Social, Universidade de Aveiro. (https://ininet.org/institutions-and-business-henry-burnay-a-case-study-maria-eugn.html)

Monteiro, Rui, Isabel Rial & Eivind Tandberg (2020) 'Fiscal risks in infrastructure', Chapter 11 in Schwartz et al. (editors) *Well Spent: How Strong Infrastructure Governance Can End Waste in Public Investment*, Washington, DC: International Monetary Fund. (https://www.elibrary.imf.org/view/IMF071/28328-9781513511818/28328-9781513511818/ch11.xml)

NAO 2018 – United Kingdom, National Audit Office (2018) *Report on PFI and PF2*, January.

PRIF 2013 – Pacific Region Infrastructure Facility (2013) *Infrastructure Maintenance in the Pacific: Challenging the Build-Neglect-Rebuild Paradigm*, Sydney. (https://www.theprif.org/sites/default/files/documents/build_neglect_rebuild_revised_full_report_2014.pdf)

Roberts, Alasdair (2010) *The Logic of Discipline: Global Capitalism and the Architecture of Government*, New York: Oxford University Press.

Saussier, Stéphane (2013) "Public-Private Partnerships", *Journal of Economic Behavior and Organization*, vol. 89, May, pp. 143–144.

Schwartz, Gerd, Manal Fouad, Torben Steen Hansen & Geneviève Verdier (editors) (2020) *Well Spent: How Strong Infrastructure Governance Can End Waste in Public Investment*, Washington, DC: International Monetary Fund. (https://www.elibrary.imf.org/view/books/071/28328-9781513511818-en/28328-9781513511818-en-book.xml)

Singh, Ram (2018) 'Does choice of procurement matter for cost and quality of infrastructure? Comparison of quality and cost of roads', *IGC Working Paper S-89209-INC-2*, London: International Growth Centre, London School of Economic and Political Science. (https://www.theigc.org/wp-content/uploads/2018/06/Singh-2018-working-paper.pdf)

TC 2001 – Portugal, Tribunal de Contas (2001) *Auditoria ao Acordo Global celebrado entre o Estado e a Lusoponte, Volume I: Introdução, Conclusões, Desenvolvimento e Observações de Auditoria* (Relatório nº 47/01 - 2ª Secção), Lisboa, November. (https://www.tcontas.pt/pt-pt/ProdutosTC/Relatorios/RelatoriosAuditoria/Documents/2001/rel047-2001-2s.pdf)

UTAP 2017 – Portugal, Unidade Técnica de Acompanhamento de Projectos (2019) *Hospital de Braga: Relatório Intercalar da Equipa de Projeto* (constituída e nomeada ao abrigo do Despacho 8300/2016, de 16 de Junho, da Coordenadora da UTAP), April 23rd. (https://www.utap.gov.pt/Publicacoes_utap/Relat%C3%83%C2%B3rio%20Fundamentado%20relativo%20ao%20Hospital%20de%20Braga.pdf)

UTAP 2019a – Portugal, Unidade Técnica de Acompanhamento de Projectos (2019) *Hospital de Vila Franca de Xira: Relatório Intercalar da Equipa de Projeto* (constituída e nomeada ao abrigo do Despacho 8323/2018, de 6 de Agosto, da Coordenadora da UTAP), May 3rd. (https://www.utap.gov.pt/Publicacoes_utap/Relat%C3%83%C2%B3rio%20Fundamentado%20relativo%20ao%20Hospital%20de%20Vila%20Franca%20de%20Xira.pdf)

UTAP 2019b – Portugal, Unidade Técnica de Acompanhamento de Projectos (2019) *Hospital de Loures: Relatório Intercalar da Equipa de Projeto* (constituída e nomeada ao abrigo do Despacho 8323/2018, de 6 de Agosto, da Coordenadora da UTAP), December 13th. (https://www.utap.gov.pt/Publicacoes_utap/Relat%C3%B3rio%20Fundamentado%20relativo%20ao%20Hospital%20de%20Loures.pdf)

Yescombe, E. R. (2002) *Principles of Project Finance*, Amsterdam: Academic Press.

Yescombe, E. R. (2007) *Public-Private Partnerships: Principles of Policy and Finance*, Amsterdam: Elsevier.

World Bank (2016) *Value for Money: Achieving VfM in Investment Projects Financed by the World Bank*, July, Washington, DC (http://pubdocs.worldbank.org/en/274711479159288956/Guidance-on-VfM-FINAL.pdf)

World Bank (2018) *Policy Guidelines for Managing Unsolicited Proposals in Infrastructure Projects*, (Washington, DC), which comes in 3 volumes:

Volume 1: Main Findings & Recommendations (https://ppp.worldbank.org/public-private-partnership/sites/ppp.worldbank.org/files/documents/UnsolicitedProposals_Volume1_MainFindings_WEB%20%281%29.pdf);

Volume 2: Guidelines for the Development of a Policy for Managing Unsolicited Proposals in Infrastructure Projects (https://ppp.worldbank.org/public-private-partnership/sites/ppp.worldbank.org/files/documents/Unsolicited Proposals_Volume2_Guidelines_WEB%20%281%29.pdf);
Volume 3: Review of Experiences with Unsolicited Proposals in Infrastructure Projects (https://ppp.worldbank.org/public-private-partnership/sites/ppp.worldbank.org/files/documents/UnsolicitedProposals_Volume3_Review_WEB%20%281%29.pdf.

10. State-owned enterprises: Struggling to be efficient[1]

Paulo Medas[2] *and Mouhamadou Sy*[3]

1. INTRODUCTION

State-owned enterprises (SOEs) are very diverse and numerous across advanced and developing economies alike. A majority are relatively small and owned by subnational governments, but some SOEs are among the largest firms in domestic economies and are owned by central governments. SOEs have become a larger part of the global economy over recent decades reflecting the growth in emerging market economies (IMF, 2020). Some are created by a specific law or a law for SOEs, while others are created under the general corporate law and quoted on the stock market. Some are essentially an agency of government, and in many cases heavily dependent on budget transfers, whereas others have a mix of public and private owners and a greater commercial focus.[4]

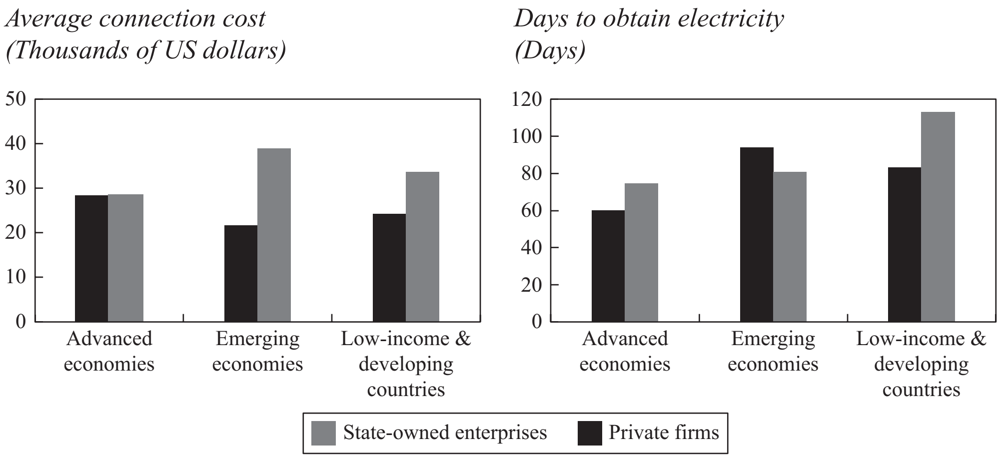
A fundamental difference with private firms is that SOEs have a public mandate set by governments. While private firms are driven by profit maximization objectives, SOEs have to balance those with public mandates that may affect their cost structure and profits. Government ownership is often justified to correct market failures. One reason is the presence of natural monopolies, especially when fixed costs are substantial and involve significant risks. Other reasons are the provision of public goods and essential services or the presence of large externalities. SOEs are also often responsible for managing governments' interests in the extractive industries. National oil companies are among the largest corporations in the world (Brazil, China, Mexico, Saudi Arabia, Russia) and control about half of the world's hydrocarbon reserves.

SOEs are prominent in network industries (water, power, transportation). SOEs in the power sector are especially dominant in transmission and distribution given that they tend to be natural monopolies. Similar factors—natural monopoly, large fixed costs and externalities—provide a rationale for government intervention in the water sector (Menard and Peeroo 2011; World Bank 2004). Delivery systems require major investments in infrastructure, and potable water and adequate sewerage are essential for public health. Most countries have opted for a high degree of public provision through SOEs. Among advanced economies, public provision of water services is dominant in the majority (for example Australia, Germany, Japan and the United States; Pérard, 2009). The provision of public transportation, including at the local level, has involved significant government intervention justified by the need to ensure affordability as well as to address congestion, pollution or accidents. Local SOEs commonly provide ground transportation in advanced economies. For example, in the United States, governments or SOEs (transit authorities) account for more than 80 percent of transit services, including buses, rail and ferries (U.S. Department of Transportation, 2019). In addition, SOEs accounted for more than half of the value all infrastructure project commitments in emerging market economies and low-income developing countries (World Bank, 2017).[5]

The presence of SOEs is also justified by governments as tools to pursue industrial policy or other development goals, including kick starting new sectors in developing countries. However, SOEs can also be found producing goods and services in competitive, mature sectors (for example soft drinks, cars, hotels or resorts) where the traditional reasons for their existence are not relevant.[6] In some cases, this is based on historical factors, including saving companies during past economic crises or being a former member of the Soviet Union. SOEs are also used to pursue broader policy goals, such as promoting employment. The growing presence of SOEs in global markets has also generated new research to understand their objectives, especially possible noncommercial objectives in foreign countries (Cuervo-Cazurra et al., 2014).

2. POLICY MANDATES AND EFFICIENCY

The performance of SOEs can have large effects on public finances, the economy and social outcomes. SOEs are given a diverse set of policy mandates by governments including provision of public goods at affordable prices, building infrastructure and promoting development goals. However, many have raised concerns that many SOEs are inefficient, pose a significant burden on government budgets and are a channel for corruption.[7] As providers of key inputs, inefficiencies in SOEs can undermine other sectors of the economy as it may lead to unreliable provision of services (for example power shortages) or higher costs (Figure 10.1). If SOEs are less efficient than private competitors, but are protected from competition, it can also



Note: Assessments based on the biggest city in an economy plus the second biggest for very large economies such as China, Japan or the United States. In total, 201 city distribution systems across 190 economies are included in the sample. An electricity distribution company is classified as an SOE if a state and/or central government owns more than 50 percent of company's equity. For the connection cost, the average is the simple average of the median cost for each country.

Sources: World Bank Doing Business and authors' calculations.

Figure 10.1 Supply of electricity

undermine productivity in the economy as a whole by leading to a misallocation of resources (Restuccia and Rogerson, 2019; Song et al., 2011).

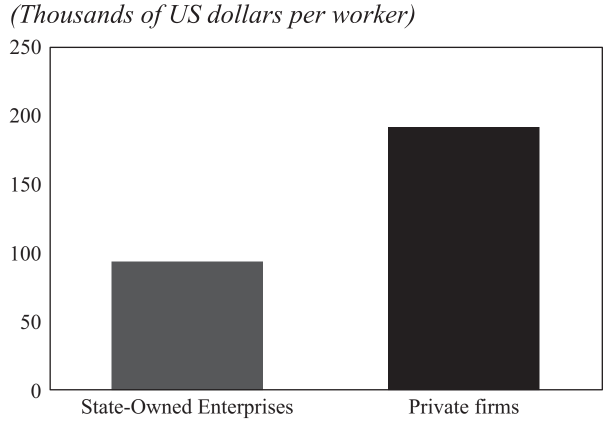
SOEs, especially in emerging market economies and low-income developing countries, have faced challenges in trying to achieve the policy mandates and pursue efficient and financially sustainable operations. The challenges can be broadly divided into three main groups. The first involves cases where the policy mandates are not appropriately funded, usually because they are not clearly specified or adequately costed. A second set of problems relates to wrong incentives. If SOEs benefit from advantages (for example bailouts or regulatory advantages) this could lead to weaker incentives to be efficient. Finally, another set of issues relates to weaknesses in transparency and governance. This includes corporate governance and the relationship between the government and management of the company.

2.1 Funding of Policy Mandates

One common problem faced by SOEs is that noncommercial activities are not well costed and budgeted, which can result in funding shortfalls and weaken the financial health of the firm. For example, SOEs may be mandated to sell goods at prices below cost, including providing public or universal service obligations (for example in the transport sector). Governments may also delay necessary funding due to financial constraints or not wanting to report higher fiscal deficits. In addition, some SOEs may be involved in quasi-fiscal activities that are not clearly defined. This includes, for example, local governments wanting to increase employment in their regions through the SOE, or asking the firms to be involved in social activities that are not part of the core activities of the company.[8] These unfunded policy mandates often result in future fiscal costs when governments are called on to recapitalize the SOEs due to losses.

These issues are especially prevalent in regulated markets, where firms may be unable to set prices or tariffs at cost-recovery levels without being properly compensated by governments. Such policies may be justified by policy goals, including to promote extensive coverage and affordability of services or to address other types of externalities (for example provision of water and sanitation has a significant impact on health). However, if the firms are not appropriately compensated it can undermine their operations over time. International experience suggests this is one of the reasons why a significant share of the population in low-income countries continues to lack access to electricity and water. In the power sector, network expansion and access relies largely on SOEs. Access to safe water also remains a challenge with more than 2 billion people estimated to lack safely managed water services, partly reflecting weak SOE performance (WHO and UNICEF, 2017; World Bank, 2004). A common problem in both power and water sectors is the failure to achieve cost recovery. For example, in a sample of developing countries in 2010–17, in 40 percent of power utilities tariffs were below operational costs and in almost all they did not cover all the costs (World Bank, 2019). Below-cost tariffs reduce the capacity to invest, lead to arrears and other short-term debt and undermine accessibility, sustainability and quality of services.[9]

Allowing SOEs (or even private operators) to charge prices that cover operational expenses, and much less investment, is also rare in local transportation especially among advanced economies. Governments struggle to find an appropriate balance of subsidies and tariffs to meet policy mandates (for example equity considerations, pollution and congestion). In the United States, revenues collected by the firm (mainly tariffs) account for just over one third of the operational costs in public transit. In Europe, the degree of subsidies varies significantly.



Note: The sample includes 98 NOCs and 1520 private firms.

Sources: Orbis; National Resource Governance Institute; and authors' estimates.

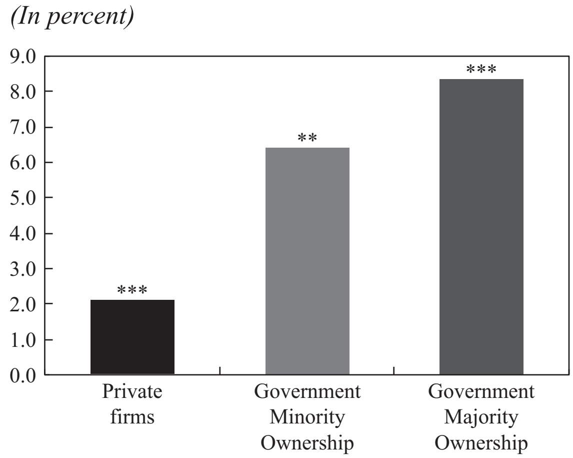
Figure 10.2 National oil companies—labor productivity

For example, for passenger rail in the European Union, revenues as a proportion of operating costs vary from less than 10 percent to more than 60 percent (Gleave, 2016). When the companies are forced to operate systematically underfunded because there is no clear mechanism to compensate for the public mandate this can lead to underinvestment and poor maintenance of the transportation infrastructure and eventually the need for large government transfers.[10]

Another obstacle for efficient operations in SOEs regards employment policies. SOEs may be constrained regarding hiring and wage policies. One recurrent example is the use of SOEs to promote job growth, leading to overemployment in the firms and higher labor costs—which then weakens the firm's efficiency and financial viability. One notable example are national oil companies (NOCs) that many oil-exporting countries create to exercise control over oil and gas exploration and garner potentially large profits for the state. However, NOCs are significantly less profitable and efficient than their private peers, partly owing to pressures from the government to engage in excessive hiring (Figure 10.2). Further evidence is that SOEs tend to create many more jobs associated with their investment. (Figure 10.3)—job intensity is found to be 30 percent higher—possibly because SOEs tend to be large and subcontract fewer jobs and that SOEs have an implicit employment remit.

2.2 Budget Constraints and Preferential Treatment

The expectation that governments will eventually compensate, or bail out, the SOE for losses may lead to the perception that SOEs face *soft budget* constraints (Kornai, 1979) and are not subject to the discipline of capital markets (Megginson and Netter, 2001). This can reduce incentives for managers to pursue efficiency gains and may lead to taking larger risks, including borrowing excessively. SOEs tend to be bailed out regularly by governments, in some cases involving large amounts (IMF, 2020; Musacchio and Pineda, 2019). In addition, SOEs



Note: The regression includes firms and year dummies. Control variables include lag of employment growth, share of oil exports as share of total exports, liquidity, solvency, total assets and sales. Standard errors are clustered at the country level. *** $p < 0.01$, **$p < 0.05$. Data are from 1999 to 2017.

Figure 10.3 Effect of 1 percent increase in net investment rate on employment

can survive longer than their private peers with systematic losses—in some sectors, more than one third of SOEs have losses for at least three years (Table 10.1). Soft budget constraints may be exacerbated by complex holding structures, which make them less transparent and difficult to monitor.

SOEs may also benefit from advantages granted by governments that shield them from competition and reduce incentives to be efficient. These could include subsidies for regular operations or access to cheaper financing with government guarantees. IMF (2020) shows that SOEs benefit from lower debt-financing costs relative to private firms for a sample of 65 countries. For example, in China, SOEs benefit from preferential access to finance, land use at below-market cost and sector-specific incentives (Bai et al., 2014; IMF, 2017; Maliszewski et al., 2016). In Vietnam, the state-owned bus company has higher operational costs than its private competitors but benefits from lower borrowing costs resulting from government guarantees

Table 10.1 Share of firms with at least three consecutive years of losses

(in percent)

	Transport	Utilities	Manufacture	Mining	Construction	Agriculture	Communication
SOEs	43.5	36.6	33.9	29.2	28.3	27.8	19.7
Private firms	12.0	15.9	10.2	14.8	8.7	12.1	10.3

Note: The measure of persistent losses is based on return on assets (ROA). Persistent losses are proxied by three years of consecutive negative ROA. Data are from 1999 to 2017.

Sources: Orbis; National Resource Governance Institute; and Baum et al. (2020).

(PPIAF, 2016). The concerns with government support also apply to other key sectors of the economy. Recent studies of the aluminum (2013–17) and semiconductor (2014–18) sectors estimated that firms, including SOEs, in these industries received government support of $70 billion and $50 billion, respectively (OECD 2019a, 2019b). In these sectors, overcapacity is an issue and there is a concern that these SOEs are less efficient than the companies that they are replacing.

2.3 Governance and Oversight

Weak governance has also proven to be a daunting challenge among SOEs. It reflects several factors. One is the lack of clear mandates and a process to ensure accountability. This is especially important when profitability is just one of the objectives, making it harder to assess the overall performance. Another problem is that in many countries, government agencies lack the capacity or information to properly monitor SOEs (Allen and Vani, 2013). In many countries, few SOEs prepare financial statements in line with international practices which are published and audited or annual reports that document their performance. In addition, lack of transparency around the operations of SOEs and the relations between the firm and government are also common. These factors make SOEs more vulnerable to corruption, including through procurement processes and hiring (IMF, 2019). In a survey, 42 percent of SOE respondents reported that corrupt acts or other irregular practices had occurred in their company during the past three years (OECD, 2018a). In addition, the evidence suggests that 80 percent of foreign bribes go to SOE officials (OECD. 2014). This is especially the case with firms operating in the natural resources sector, where economic rents can be substantial.

Many countries also either lack key elements of good corporate governance for SOEs in their laws or do not fully implement such elements in practice. The governance responsibilities of the state include proper exercise of its ownership duties—including monitoring performance regularly and avoiding undue political interference and conflicts of interest. One challenge has been transparently selecting SOE boards that are independent and qualified. For example, a study of local public utilities in Italy found that when boards were dominated by politically connected directors, SOE employment was higher and firm performance was worse (Menozzi et al., 2012). Musacchio and Pineda (2019) also note that a large number of executive directors of Latin America's SOEs lack a background in management or expertise on the sector where the company operates. Another challenge is to fully integrate good corporate governance practices in day-to-day activities, including effective internal controls and risk management systems. Even countries that were perceived to have relatively good monitoring and reporting of SOE activities have struggled with corruption in some of their largest companies (for example Brazil, South Africa).

3. CROSS-COUNTRY EVIDENCE ON PERFORMANCE

Assessing the performance of SOEs and understanding its main drivers is critical, but difficult in practice. This is complicated by the very diverse goals. Many still strive to be profitable, but also need to contend with not always well-defined or transparent public mandates that may impact profitability and their cost structure. In many instances, SOEs do not submit regularly and timely information on their activities either to the public or even the government in some

Table 10.2 SOEs' financial performance relative to private firms

	Return on Equity	Profits and losses margin	Cost per employee	Productivity	Value added per employee
	(in percent)			thousands of USD	
SOEs	0.6	1.6	31.0	19.3	37.8
Private firms	8.6	3.7	25.5	117.7	48.8

Note: Productivity is proxied by sales per employee; The nominal values are deflated using 2017 prices. Data are from 1999 to 2017.

Sources: Orbis, NRGI and authors' calculations.

cases. As such, analysis of the firms' financial performance has been very limited. In addition, most attempts in the literature have focused on country studies, and in some cases a small group of countries, with little cross-country analysis that could shed light on what drives the differences in performance across sectors and countries.[11]

This section analyzes SOEs' financial performance, and its main determinants, across more than 100 countries. The sample of close to 1 million individual firms, both SOEs and private firms over the period of 1999–2017 allows us to compare SOEs across countries and sectors and relative to private firms (see Appendix 1).[12] Importantly, the data allow us to study if there is a systematic productivity gap between SOEs and private firms and whether this depends on different country and sector characteristics.[13] Benchmarking SOEs with private firms helps to better understand their relative performance.

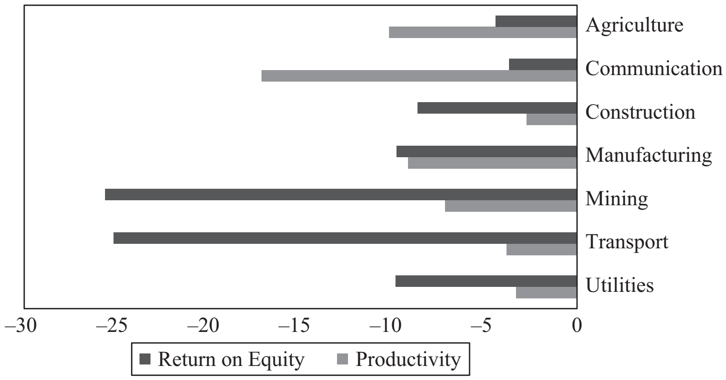
A simple comparison of financial indicators shows that SOEs perform relatively worse compared to private firms. This is true across many different indicators of financial performance. Table 10.2 shows that the return on equity among private firms is 8 percentage points higher and labor productivity is six times higher than those of SOEs.[14] The large differences in efficiency can have economy-wide implications, especially in countries where SOEs have a large presence or operate in key sectors of the economy. Even if SOEs are technically efficient, they may be unprofitable if they cannot adjust their prices to cover costs. This can put a burden on government budgets and exert a drag on growth by crowding out expenditure on areas like health and education (Shirley, 1995).

The difference in performance between SOEs and private firms could be partly explained by costs associated with SOEs' policy mandates.[15] However, other factors are also likely to be present. We analyze three main explanations discussed in the literature: (i) the role of market structure by looking at different sectors in which SOEs operate; (ii) the degree of state ownership; and (iii) the quality of governance. While we analyze the impact of each factor separately, their effect on firms' performance may differ depending on the interaction of the different factors.

3.1 Key Drivers of Performance

One question is whether what matters is the ownership structure of the firm or the sector where they operate. If SOEs operate mainly in regulated sectors such as utilities this could explain why they underperform with respect to private firms—as in these sectors they have restrictions

(Percentage point difference of SOEs relative to private firms)



Note: Productivity is proxied by sales per employee; The nominal values are deflated using 2017 prices. Data are from 1999 to 2017.

Sources: Orbis, NRGI and authors' calculations.

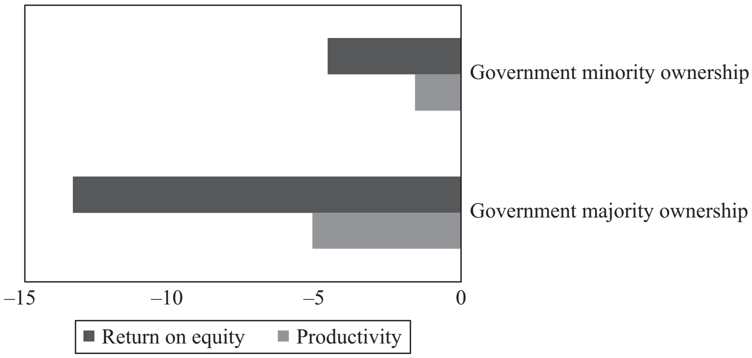
Figure 10.4 Relative performance of SOEs by sector

on setting prices and may have additional requirements or mandates that raise costs. This is similar to the argument that SOEs do not perform significantly differently from private firms if they operate in competitive environment or in the same industry (Kole and Mulherin, 1997). However, the data show that there is a gap between SOEs and private firms across all sectors (Figure 10.4), although it suggests the productivity gap may be lower in more regulated sectors—between 3 and 4 percentage points in public utilities and transportation—and larger in more competitive markets, for example the gap is 9 percentage points in manufacturing.[16]

Another question is whether the degree of government ownership matters. If governments have private partners, it may help improve performance as private investors will put more emphasis on profits and efficiency (Vining and Boardman, 1992). Simple descriptive statistics show that the degree of state ownership appears to matter. Figure 10.5 shows that private firms perform significantly better than firms where the government has a minority ownership. In turn, these perform better than firms with government majority ownership. This relationship is robust when controlling for country and sector differences. For example, private firms and SOEs with government ownership below 50 percent have returns on equity that are higher than those of SOEs that are majority owned by the government, 6 percentage points and 0.7 percentage points, respectively (Table A.3, Appendix 2).[17,18]

Finally, we examine the impact of governance on performance. As discussed, when assessing governance, it is crucial to take into account the relationship between governments and the company. Weak institutions in local or national governments are likely to impact more SOEs given the close relationship with public officials. This could include several forms of corruption, such as: (i) political influence and favoritism; (ii) poorly managed conflicts of interest; and (iii) lack of independent and professional boards and management (see Baum et al. 2019). As such, to study the impact of the quality of governance, we take a widely used indicator on the degree of control of corruption in government.[19]

(Percentage point difference relative to private firms)



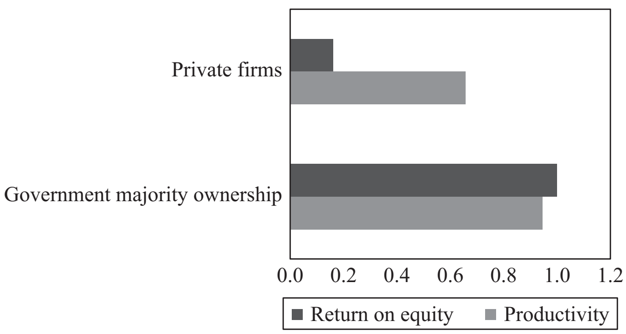
Note: Productivity is proxied by sales per employee; The nominal values are deflated using 2017 prices. Data are from 1999 to 2017.

Sources: Orbis, NRGI and authors' calculations.

Figure 10.5 Degree of state ownership and firms' performance

The evidence shows that, on average, firms in strong governance countries have better performance than in countries with low governance (Figure 10.6). This result is confirmed by cross-country analysis that controls for other factors. Table A.4 in Appendix 2 shows the results for SOEs with majority government ownership. The estimated coefficient for governance always has the expected sign and is strongly significant. Weak governance (or high corruption) is associated with lower profitability, lower productivity and value added per employee, and higher labor costs.

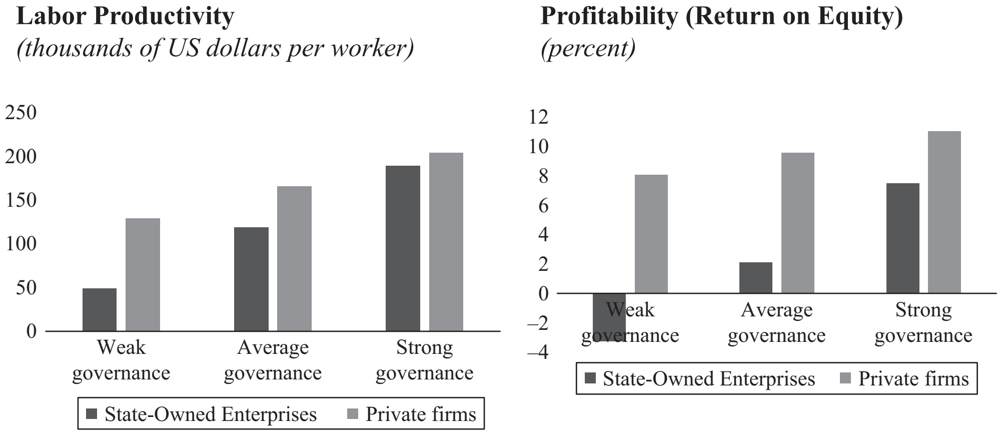
(Difference between firms in countries with strong governance and firms in countries with weak governance; percentage points)



Note: Productivity is proxied by sales per employee; The nominal values are deflated using 2017 prices. Data are from 1999 to 2017.

Sources: Orbis, NRGI and authors' calculations.

Figure 10.6 Governance and performance



Note: The panels illustrate the effect of control of corruption on firms' performance depending on the type of ownership. SOEs are firms for which the government owns 50 percent or more. The analysis controls for firm-specific characteristics, country-specific variables and sector where the firm operates. Data are from 1999 to 2017. The charts show the simulations when control of corruption is weak, average and strong .

Sources: ORBIS, Worldwide Governance Indicators, IMF (2020).

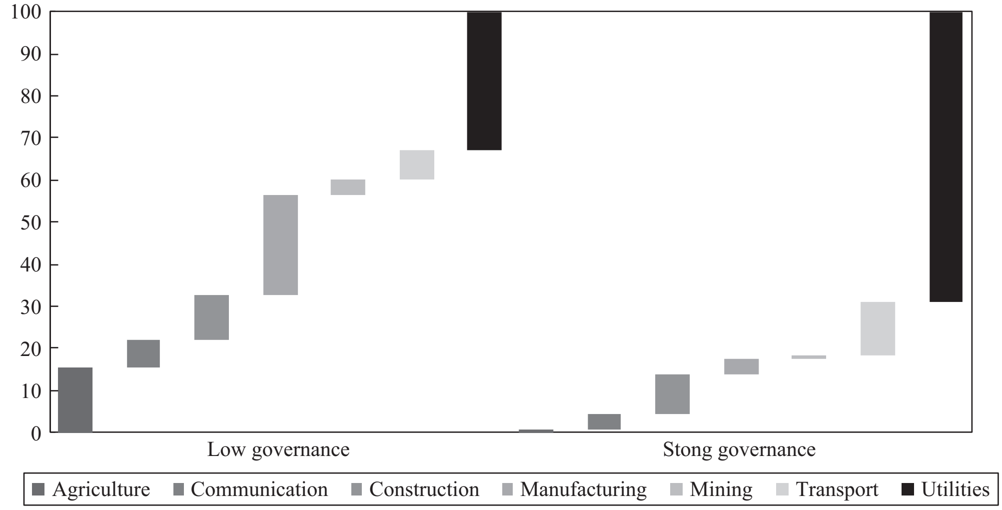
Figure 10.7 Governance, ownership and performance

How does the impact of governance depend on the type of ownership? Table A.5 in Appendix 2 presents the results. It shows that the interaction term between the control of corruption and ownership is negative for profits and productivity—that is, changes in the degree of governance will have a larger impact on SOEs than private firms. It confirms that SOEs are more vulnerable to weaker governance (corruption) in government. Figure 10.7 illustrates two key messages from the analysis. First, SOEs are significantly more profitable and efficient in countries with low corruption. Second, the performance gap between private firms and SOEs can be also explained by governance. Namely, private firms in weak governance countries are almost three times more productive than SOEs on average. That gap almost disappears in countries with strong governance (we find similar trends for profitability). The remaining (much smaller) gap may be due to the cost of public mandates.

One possible driver of performance across different degrees of governance is the sector in which the SOE operates. Countries with better governance are more selective, having SOEs in specific sectors, especially utilities and transportation, in which there is a stronger reason for intervention and the performance of SOEs is closer to that of private firms (Figure 10.8). These countries have fewer SOEs in sectors where private firms have significantly superior performance (for example, manufacturing).

4. INSTITUTIONAL FRAMEWORKS TO IMPROVE PERFORMANCE AND ACCOUNTABILITY

The international experience provides important lessons on how to design institutions and frameworks to bolster the performance of SOEs. In particular, successful countries demand that SOEs deliver value for money, that is: achieve their policy mandates (noncommercial



Note: The chart shows the share of SOEs by sector in countries with low governance and countries with strong governance. The low governance group are countries with a governance score that is below the average governance score in the sample minus one standard deviation. The strong governance group are countries with a governance score that is above the average governance score in the sample plus one standard deviation.

Sources: ORBIS, Worldwide Governance Indicators, and authors' calculations.

Figure 10.8 Share of SOEs by sector and quality of governance

objectives) efficiently and be profitable or, at least, not be a burden on the government budget. The evidence also suggests that reforms can have large impacts on profitability and efficiency of SOEs, especially reforms on pricing and SOE governance (Baum et al., 2019; IMF, 2020). The experiences also show that there is not one specific institutional feature that assures success, but instead there are mutually reinforcing elements. A more critical aspect is to ensure actual implementation of the legal and regulatory frameworks for SOEs.

4.1 Objectives, Incentives and Monitoring

An important pillar of a sound framework is that the decision on whether to have an SOE should be assessed vis-à-vis other types of government intervention or private sector participation. For example, a cost–benefit analysis may show that other alternatives may be better (or worse), including regulation with private sector participation, or some other form of public–private partnership. In addition, when using SOEs to achieve broader policy goals, such as reducing poverty or generating jobs, more traditional fiscal policy instruments could be more efficient (for example targeted transfers to poorer households).

If using SOEs, governments need to ensure their boards and managers have the right incentives to run the firm efficiently and with sound risk management in order to deliver value for money. This can be particularly challenging given the difficulty in observing management efforts in SOEs that are not run solely on a commercial basis and the complexity in setting measurable performance objectives (for example quality of service or social goals). A robust framework needs to (i) have clearly defined and specific goals, both regarding policy mandates (for example economic, social, environmental) and financial performance; and (ii) ensure managers of SOEs have the right incentives to pursue those objectives. This will also help avoid using the company for less transparent activities and that may carry substantial costs that have not been properly vetted or budgeted.

Governments should set clear ownership objectives that are regularly reviewed to remain relevant and that the reasons for the existence of the SOE remains valid. For example, the Nordic countries in Europe specify their objectives and review their framework, including the rationale for ownership or changes in ownership policy, on a regular basis (OECD, 2018b). Ownership policies should clearly state (i) the mandates, objectives and a dividend policy for SOEs; (ii) the approach to achieving professional boards of directors; (iii) the functions carried out by the government as owner of the SOE and its coordination with fiscal risk oversight functions; and (iv) the way the government exercises its ownership rights (Allen and Alves, 2016).

One of the thorniest aspects is to clearly identify noncommercial objectives and ensure consistency between them and the financial targets (Box 10.1). It may be difficult to precisely define the noncommercial objectives, but it will be important to set targets that are as specific as possible, cost them, and define a compensation if they imply large losses to the company.[20] Some countries (for example Australia) require SOEs to add value to their shareholders by gaining efficiency in operation and pricing, thereby earning at least a commercial return, which would help ensure the financial sustainability of the business. This may require that the government partly (or fully) funds the costs of noncommercial activities through subsidies or other transfers. Some countries adjust the targeted profitability ratios by uncompensated costs of the noncommercial activities. In Sweden policy targets and financial targets are normally set together, which allows the public policy targets to be balanced against the impact on financial returns.

BOX 10.1. POLICY MANDATES IN REGULATED MARKETS

The tension between achieving policy goals and ensuring the SOE is financially sound and efficient is particularly pronounced in regulated markets. In sectors such as public utilities and transportation, governments are usually trying to balance many interests including achieving universal access at affordable prices and normal returns on investments whilst taking into account externalities associated with these activities—for example, impact on health and education of having poor water and electricity services or pollution from traffic. Some of the issues are relevant whether the monopolist operator is private or state owned (for example incentive to overcharge consumers and underinvest); however, some others are specific to SOEs. This includes the firms being given unfunded mandates that eventually lead to poorer services, low efficiency, and eventually create fiscal risks to the budgets. Conversely, SOEs may be given preferential advantages (funding, regulatory) over private firms that are more efficient.

Getting the price right. Preferably, prices should be set to ensure cost recovery (including reasonable return on new investments). However, this is not always straightforward and there may be complex design issues.

- Prices may be set below cost, especially in low-income countries, where a large segment of the population is poor and social safety nets are not well developed. In such cases, there should be a transparent process for governments to provide appropriate compensation in a timely manner that ensures the firm's financial sustainability and incentives for managers to be efficient (for example, compensation could be based on cost benchmarks). Similarly, it is important to prevent excessively high prices if the SOE has monopoly power because high prices may lead to inefficiencies.

- The pricing policy in sectors with negative externalities (for example, fossil fuels that lead to pollution and health problems) should also be adjusted. This may undermine other goals, including equity, and as such needs to be accompanied by other actions, such as support to poorer households.
- Pricing decisions should be transparent and depoliticized, for example by introducing automatic adjustment mechanisms for fuel prices or tariffs.

Independent regulatory agencies need to balance different interests, ensuring that government and firms operate according to transparent and well-defined rules. For example, regulators can ensure tariffs in public utilities are set to balance affordability with the need to cover costs. Regulators can also promote a level playing field when private investors are involved. When dealing with SOEs, regulators can help by encouraging reforms to improve efficiency (including governance), and setting benchmarks for costs and quality indicators, and assess whether revenues are enough to cover for needed investment and maintenance expenditures. However, when SOEs are monopolists or have a dominant role, the effectiveness of regulators has been much lower than when there are private competitors—possibly because SOEs feel more constrained to follow regulations (World Bank, 2019). Another challenge is to develop enough capacity in regulatory agencies to analyze complex issues, such as setting appropriate pricing policies when there are conflicting objectives and limited information.

Governments also need to develop instruments to ensure management incentives are aligned with the owner's priorities. SOE management should have the autonomy to decide how best to pursue the objectives but needs to ensure they are in line with the policy priorities. This requires striking the right balance between effective oversight and limiting (undue) political interference. In Sweden and Norway, the government uses the owner's dialogue—regular meetings between the owner and the company—to track financial and public policy targets. In Norway, the ownership policy clarifies that opinions conveyed by the state during these meetings are suggestions and the board makes the decisions. The Danish ownership policy also qualifies that government communication with company management must not imply that the minister de facto leads the corporation. In practice this balance it not always easy to achieve especially if governance is weak.

The monitoring and regular assessment of SOE management can be done through performance contracts, but their effectiveness will depend on their design and on supportive institutions. They need to contain a clear definition of the objectives, strategy, a set of performance indicators, reporting obligations, the dividend policy and a clear allocation of responsibilities for non-compliance. Performance contracts can be complex to design and difficult to implement, especially in setting both the commercial and noncommercial obligations—they need to be clear on these targets and consistent with broader ownership policy objectives. But these contracts are not a panacea. They have been relatively effective in some countries, especially those with high governance overall. For example, New Zealand has a sound and effective performance contracting framework within which SOEs' goals are informed by risk oversight and fiscal objectives. On the other hand, Shirley (1998) and Simpson and Nyante (2015) found that the implementation of performance contracts among developing countries has not been effective. Weaknesses have been associated with (i) multiple or changing objectives or lack of link between targets and the effort of the manager; (ii) unrealistic objectives; (iii) ill-designed incentives for management (for example bonus, sanctions); (iii) insufficient monitoring, often due to weak involvement by the board in the design and execution of the agreements, or to a faulty assessment of the costs of noncommercial activities; and (iv) lack of legal mechanisms or political will to enforce the contracts.

Another frequent problem is that monitoring is underdeveloped. In some countries, governments do not even have individual or aggregate comprehensive information on all SOEs they own. For effective monitoring, all SOEs should be required to provide timely, regular and reliable reports on their operations. If there are capacity constraints, monitoring could prioritize a subset of SOEs in a first stage. For example, those that run complex business models (where it is difficult to assess performance and risks), represent large fiscal risks or SOEs where disruptions could have negative economic effects. Governments should also strive to benchmark SOEs' financial performance to other SOEs (and private firms) in other countries and in the same sectors. This can provide an indication of potential weaknesses if an SOE performs worse than its sectoral peers in other countries. The benchmarks can be useful to help set goals and expectations regarding the financial performance of the SOE.[21] Governments should also establish sound systems for monitoring potential fiscal risks and costs (Baum et al. 2021) in relation to direct and indirect transactions with SOEs (repayment of budget loans, potential for call on guarantees, lower than anticipated dividends).

The degree of transparency regarding SOEs also remains limited, preventing greater external scrutiny by the public and markets. Financial information on SOEs remains very uneven and in many countries sparse. This is especially the case for national oil companies, which should require a higher degree of transparency given they manage large assets of the country (NRGI, 2019). Disclosure of SOE financial statements is the prevailing practice in advanced economies, whereas in emerging market economies disclosure is often restricted to listed SOEs. Publishing regular reports with detailed information and analysis of the performance of the SOE sector at the aggregate and company levels is an important step towards greater transparency and accountability.[22]

Countries also need to develop institutional capacity to effectively monitor SOEs. One of the problems is that a fragmented approach, with several agencies and ministers sharing responsibility, can weaken the ability to do proper oversight. A strong oversight and control agency can yield better performance from SOEs (Musacchio and Pineda, 2019). A centralized model provides the potential for ensuring consistency between the ownership (for example, representation on company boards, strategic direction of firms) and financial oversight functions. It could take the form of an autonomous agency or holding company (as in Finland, France, Kenya, Malaysia, Peru and Singapore). Holding companies—particularly in countries with sound governance and oversight arrangements in place—exhibit advantages when managers have professional expertise and they protect SOEs from undue political interference. Whatever the specific model, however, having one government unit responsible for the financial oversight of SOEs (for example in the Ministry of Finance) has important benefits including making oversight activities more coherent, while pooling experts from different areas. In addition, especially for large SOEs with complex businesses, bringing in private shareholders (mixed ownership) could help promote greater focus on efficiency and provide another layer of external oversight.

4.2 Corporate Governance

Another pillar of the framework is establishing and enforcing SOE good corporate governance standards. The OECD Guidelines on Corporate Governance of SOEs (OECD, 2015) is the most well-known benchmark in this area. The composition of SOE boards plays a significant role in the quality of corporate governance. There has been a growing call for promoting independent and professional boards that can help ensure proper accountability. In some countries, some or all of the members of the boards of directors are required to be independent of the government (for example, Canada, Germany, Netherlands and Switzerland). However, implementing high corporate governance standards remains challenging in many low-income and developing countries.

Appropriate regulation of SOEs is another important element of corporate governance. In some countries, at least the largest SOEs are subject to the same regulatory framework as listed private companies (for example Chile, Netherlands, Norway and Sweden). Adoption of bankruptcy laws or non-bailout clauses can be useful to impose discipline but may not be enough in practice. It requires a high degree of transparency, namely comprehensive and frequently disclosed financial information of SOEs. In addition, SOEs should not have preferential access to other financing sources, such as cheap credit from public banks or preferential loans from other non-financial SOEs. Moreover, the strength of incentives will also depend on the government's commitment to a no-bailout clause (or allow a SOE to go bankrupt). It may be difficult or unrealistic that governments will not support SOEs that have a systemic importance (for example, could have a significant detrimental impact on economic activity) or provide critical basic services (for example water) such a commitment is like to be less credible. In such situations, financial support should be accompanied by plans to strengthen the business model of the company and address the reasons behind the financial problems.

4.3 A Comprehensive View of Public Finances

Given that SOEs use public resources and pursue policy goals, it is important to ensure that they operate consistently with the broader fiscal objectives. However, this is an area that is much less developed. In general, SOEs are not consolidated in fiscal statistics and discussion in government budgets tends to be limited. The lack of comprehensive data and policy strategy has important disadvantages. First, it means there is no competition for the best use of public resources. SOEs may receive funds from the budget that could be better and more efficiently used in other areas or priorities. Second, it may undermine a full diagnostic of potential vulnerabilities of the public sector—for example, management of assets and liabilities will be hampered if it does not take into consideration SOEs' balance sheets, which may have large assets or be highly indebted. Third, it reduces transparency on the financial interactions between the government and SOEs. In recent years, there is greater awareness of the need for progress in these issues.

There is a growing trend to gradually expand the coverage of fiscal reporting towards the whole public sector. Some countries or regions already implement a public sector balance sheet approach (Australia, New Zealand, United Kingdom) or partially reflect SOEs in the public accounts (for example, common in Latin America). But for many others data coverage beyond budgetary central government remains very limited.[23] It will be important to make further progress towards integrating SOEs into the fiscal accounts and in the public sector balance sheet.[24] One widespread weakness is that quasi-fiscal (noncommercial) activities and non-monetary transfers are generally not appropriately reflected in government budgets.[25] More efforts are needed to transparently disclose government mandates in the budget, including the costs and fiscal risks associated with the operations of the SOEs.

There is also a about debate whether to include non-financial SOEs in fiscal targets (for example deficits and debt). Doing so, would create greater incentives for fiscal discipline and transparency because governments will likely exercise greater oversight over SOEs' overall borrowing and the ability to circumvent fiscal targets would be more limited. Inclusion would ensure that the broader fiscal policy goals are consistent across the public sector. Many governments in Latin America already include most non-financial SOEs in the fiscal targets; but it is unusual in other regions of the world. Governments should ensure comprehensive coverage in fiscal targets of at a minimum non-financial SOEs that pose significant fiscal risks and for which the government is a majority shareholder.

APPENDIX 1: DATA AND EMPIRICAL METHODOLOGY

A1.1 Data on Firms

We use the Orbis database as our primary source. SOEs in Orbis are identified through ownership as ‘organizations ultimately owned or de facto controlled by the public sector entities’. The analysis is based on unconsolidated financial data of domestically owned SOEs in sectors, excluding the financial sector. Our secondary source for data on national oil companies (NOCs) comes from the Natural Resource Governance Institute. This database covers 71 NOCs headquartered in 61 countries. Finally, for some countries, we complement the data from national authorities and data collected by IMF staff. The final sample contains about 15 000 majority state-owned enterprises and 4000 minority state-owned enterprises. Table A.1 in Appendix 2 contains the list of 102 countries in the SOEs sample between 1999 and 2017 after cleaning the data (following the approach in Baum et al. 2019).

A1.2 Econometric Approach

How does the degree of state ownership affect the performance of SOEs? To what extend does governance and sector affect the performance of SOEs compared to private firms? Formally, the relationship between the performance of SOEs and governance or ownership can be described as:

$$PER_i = \alpha_0 + \alpha_1 Ownership / Governance + \alpha_2 X_{i,t} + \varepsilon_i \qquad (10.1)$$

where $PER_{i,t}$ represents a specific performance indicator of an SOE i at time t. $X_{i,t}$ is a vector of observable covariates and ε_i is a vector of unobservables. Many variables are used to gauge the performance of SOEs. Some variables assess their profitability (return on equity and operating margin) and others to assess their efficiency (cost of employees per operating revenue, sales per employee, and value added per employee). The variables of interest are ownership and governance. Ownership varies across firms from 0 percent (private ownership) to 100 percent (full government ownership) and is fixed over time due to data limitations (and being slow changing in time). Given that there is no database with a systematic measure of governance at the firm level, we use a measure of governance at the country level. The estimation uses the control of corruption (CoC) index from the World Governance Indicators (WGI) as a proxy of governance. Governance measures—including the CoC of the WGI—tend to be highly persistent (almost time invariant). Therefore, estimating equation (10.1) using cross-country standard regression techniques such as the fixed-effect (FE) estimation to isolate the effect of governance is challenging as the FE transformation eliminates all time-invariant regressors. To tackle this issue, we use the classical pooled-cross section regression model while considering the heteroskedasticity in the sample and differences in performance between firms due to the fact that they are operating in different sectors.

To analyze the different impacts of governance on SOEs and private firms, we expand equation (10.1) by building on a similar approach as Dewenter and Malatesta (2001), who focus on the effect of ownership on firm performance:

$$PER_i = \alpha_0 + \alpha_1 Ownership_i + \alpha_2 Ownership_i G_k + \alpha_3 G_k + \alpha_4 X_i + \alpha_5 Y_k + {}_i + \varepsilon_i \qquad (10.2)$$

$Ownership_i$ is a dummy variable that allows us to identify private firms and SOEs. In some specifications, we use the exact value of ownership instead of the dummy. G_k is a measure of the quality of governance in country k. X_i represents a set of firm-level characteristics such as liquidity, solvency and other firm-level characteristics suggested in the literature. μ_i are sectoral dummies. $Y_{k,t}$ represents some non-firm level controls such as real GDP growth, GDP per capita (PPP) and natural resource endowment. We control for GDP per capita as the performance of firms may be correlated with the level of development. We control for natural resource endowment—proxied as the share of oil exports to total exports—because it can affect both the performance of firms and corruption. Indeed, the literature (see for example Brollo et al., 2013) suggests that windfalls associated with natural resources may exacerbate corruption, while at the same time raising the profitability of firms, particularly in the extractive sectors. We also control for a dummy for transition economies from the former Soviet Union given the importance of SOEs in these countries.

A1.3 Robustness Checks

We ran a number of exercises to test the robustness of the results. First, the analysis used the control of corruption indicator as a proxy for a country's governance. As an alternative, we use a government effectiveness indicator (Kaufmann et al., 2007, 2010).[26] Second, because the coverage in Orbis varies significantly across countries, we keep only countries with a coverage of at least 60 percent.[27] A third robustness check is done to improve the comparison between SOEs and private-owned firms (POEs) in the sample. Most of the companies in the sample—about 98 percent—are private firms. To make the comparison meaningful, we restrict the sample such that we only keep private firms within a sector that has at least one SOE. Fourth, to address potential sample bias, we restrict countries in the sample of POEs to be the same as those in the sample of SOEs. If firms in the private sector primarily belong to countries with stronger governance and those in the public sector are in countries with weaker governance, it could bias the results towards the conclusion that weak governance hinders SOEs more than POEs.

We also use an alternative two-step approach—instead of a pooled regression—to deal with the fact that some of our variables are time-variant (for example leverage and GDP growth) while others are time-invariant (ownership) or almost time-invariant (governance). Hsiao (2003) was the first to propose the two-step approach. In the first step, equation (10.1) is estimated by using the within estimator (fixed effects) and including only time-varying regressors. The first step allows us to include firm-fixed effects to identify the effect of all characteristics that are time-variant, introduce lagged dependent variables, and use clustered standard errors. Given that the first step includes FEs, the FE transformation eliminates all time-invariant or almost time-invariant regressions. This is why in the second step the estimated fixed effects are regressed on slowly moving variables (for example governance) with a between estimator.[28] This allows us to identify their effects.

In all these robustness checks, the results are similar to those of the baseline regressions both qualitatively and quantitatively.[29]

APPENDIX 2: TABLES

Table 10.A1 Distribution of countries in the SOE sample by income level

Advanced Economies	Emerging Markets		Low-Income Countries
Austria	Algeria	Macedonia, FYR	Bangladesh
Belgium	Angola	Mexico	Cameroon
Cyprus	Argentina	Montenegro	Chad
Czech Republic	Azerbaijan	Namibia	Democratic Republic of Congo
Denmark	Bahrain	Oman	Republic of Congo
Estonia	Bolivia	Pakistan	Cote d'Ivoire
Finland	Bosnia & Herzegovina	Panama	Ethiopia
France	Brazil	Peru	Ghana
Germany	Brunei Darussalam	Philippines	Kenya
Greece	Bulgaria	Poland	Liberia
Iceland	Cabo Verde	Qatar	Moldova
Italy	Chile	Romania	Mozambique
Japan	China	Russia	Myanmar
Korea	Colombia	Saudi Arabia	Nepal
Latvia	Croatia	Serbia	Nigeria
Lithuania	Ecuador	South Africa	South Sudan
Luxembourg	Egypt	Suriname	Sudan
Malta	Equatorial Guinea	Thailand	Tanzania
Netherlands	Gabon	Trinidad and Tobago	Timor-Leste
Norway	Hungary	Tunisia	Yemen
Portugal	India	Turkmenistan	
Singapore	Indonesia	Ukraine	
Slovakia	Iran, I. Rep. of	United Arab Emirates	
Slovenia	Iraq	Vietnam	
Spain	Jamaica	Venezuela	
Sweden	Kazakhstan		
Switzerland	Kosovo		
Taiwan: Province of China	Libya		
28 countries	**53 countries**		**21 countries**

Source: Authors' compilation.

Table 10.A2 The effect of ownership on SOE performance

	Return on Equity	Return on Asset	Operating Profit per Sales	Labor Costs per Op. Revenue	Productivity: Sales per Employee	Value Added per Empl.
	[1]	[2]	[3]	[4]	[5]	[6]
GDP growth	0.1567***	0.0549***	0.1773	−0.7807***	1.8472***	−2.0883***
	(0.0379)	(0.0179)	(0.1279)	(0.0864)	(0.4069)	(0.4125)
Share of oil exports as a share of total exports	−0.0511***	−0.0210***	−0.1109***	0.3938***	−2.1337***	−0.7403***
	(0.0125)	(0.0059)	(0.0420)	(0.0300)	(0.1332)	(0.1434)
Liquidity: current ratio	0.0013***	0.0007***	0.0025***	−0.0006*	0.0099***	0.0096***
	(0.0001)	(0.0001)	(0.0005)	(0.0003)	(0.0014)	(0.0017)
Solvency (Shareholders funds/Total assets)	−0.0112***	0.0198***	−0.0273***	0.0823***	−0.6434***	−0.0572*
	(0.0030)	(0.0014)	(0.0101)	(0.0068)	(0.0313)	(0.0319)
Total employment	0.0040***	0.0027***	0.0175***			
	(0.0004)	(0.0002)	(0.0013)			
Sales	0.0323***	0.0168***	0.0641***			−0.0001
	(0.0007)	(0.0004)	(0.0025)			(0.0101)
Total assets					0.2008***	0.1918***
					(0.0031)	(0.0041)
GDP per capita	0.0062***	0.0050***	0.0204***	−0.0355***	0.8985***	1.2659***
	(0.0015)	(0.0007)	(0.0050)	(0.0036)	(0.0160)	(0.0165)

Table 10.A2 (Continued)

	Return on Equity	Return on Asset	Operating Profit per Sales	Labor Costs per Op. Revenue	Productivity: Sales per Employee	Value Added per Empl.
	[1]	[2]	[3]	[4]	[5]	[6]
Transition economies (dummy)	−0.0297***	−0.0155***	−0.0748***	0.0408***	−0.7155***	–0.3327***
	(0.0024)	(0.0012)	(0.0082)	(0.0055)	(0.0263)	(0.0276)
Ease of starting a business	0.0906***	0.0394***	0.1465***	−0.1034***	0.7979***	0.1723**
	(0.0067)	(0.0032)	(0.0225)	(0.0158)	(0.0716)	(0.0686)
Exact ownership	−0.0076***	−0.0053***	−0.0152**	0.1108***	−0.3103***	−0.1651***
	(0.0019)	(0.0009)	(0.0065)	(0.0045)	(0.0204)	(0.0195)
Industry fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.4272***	−0.2218***	−0.9107***	0.9722***	–14.5098***	–16.4764***
	(0.0322)	(0.0152)	(0.1081)	(0.0705)	(0.3387)	(0.3098)
Observations	110 692	108 678	108 434	87 571	110,160	44,120
R^2	0.2126	0.2380	0.1097	0.3250	0.8117	0.8636
Number of firms	16 224	16 142	15 963	13 710	16,123	7,841

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: Standard errors in parentheses. Weighted least squares are used to correct for heteroskedasticity. *** p< 0.01, **p<0.05, *p<0.1. The time dimension ranges from 1999 to 2017. Productivity and value added per capita are in logarithms. The Orbis database provides information about the share of the state ownership in SOEs. This is called 'exact ownership' in the table. In Table A.3., the exact ownership is used to create a dummy for government majority ownership (ownership between 50 percent and 100 percent), government minority ownership (ownership below 50 percent) and private firms (government ownership equals 0 percent).

Table 10.A3 The relative performance of SOEs with government majority ownership

	Return on Equity	Return on Asset	Operating Profit per Sales	Labor Costs per Op. Revenue	Productivity: Sales per Employee	Value Added per Empl.
	[1]	[2]	[3]	[4]	[5]	[6]
GDP growth	0.6092***	0.2862***	0.3730***	−0.4885***	5.9720***	0.1013***
	(0.0061)	(0.0025)	(0.0052)	(0.0139)	(0.0426)	(0.0068)
Share of oil exports as a share of total exports	−0.0381***	−0.0122***	0.0093***	0.2716***	−1.6999***	0.0185***
	(0.0027)	(0.0011)	(0.0023)	(0.0048)	(0.0195)	(0.0024)
Liquidity: current ratio	0.0007***	0.0000***	0.0013***	−0.0020***	−0.0084***	0.0006***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0000)
Solvency (Shareholders funds/Total assets)	−0.0302***	0.0796***	0.0264***	0.1060***	−0.1851***	0.0234***
	(0.0005)	(0.0002)	(0.0005)	(0.0009)	(0.0039)	(0.0004)
Total employment	0.0021***	0.0006***	0.0024***			
	(0.0001)	(0.0000)	(0.0001)			
Sales	0.0262***	0.0117***	−0.0087***			-0.0027***
	(0.0001)	(0.0001)	(0.0001)			(0.0001)
Total assets					0.2618***	0.0123***
					(0.0005)	(0.0001)
GDP per capita	−0.0108***	−0.0041***	−0.0036***	0.0667***	0.9574***	0.0474***
	(0.0002)	(0.0001)	(0.0002)	(0.0004)	(0.0017)	(0.0002)
Ease of starting a business	0.1778***	0.0861***	0.0983***	−0.0948***	0.6012***	0.0316***
	(0.0010)	(0.0004)	(0.0008)	(0.0016)	(0.0070)	(0.0007)

Table 10.A3 (Continued)

	Return on Equity	Return on Asset	Operating Profit per Sales	Labor Costs per Op. Revenue	Productivity: Sales per Employee	Value Added per Empl.
	[1]	[2]	[3]	[4]	[5]	[6]
SOEs with ownership below 50%	0.0068***	0.0093***	0.0160***	−0.1125***	0.2223***	0.0122***
	(0.0017)	(0.0007)	(0.0015)	(0.0029)	(−0.0124)	(0.0001)
POEs	0.0600***	0.0384***	0.1072***	−0.1482***	0.8454***	0.0081***
	(0.0009)	(0.0004)	(0.0008)	(0.0015)	(0.0066)	(0.0008)
Industry fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.6674***	−0.3876***	−0.4471***	0.0969***	−15.2013***	-0.5680***
	(0.0044)	(0.0018)	(0.0038)	(0.0069)	(−0.0308)	(0.0037)
Observations	4 127 834	4 051 078	4 035 089	3 571 875	40,87,402	2,477,466
R^2	0.1217	0.2567	0.0778	0.0846	0.5156	0.2240
Number of firms	862 393	856 094	850 391	720 218	8,58,854	500,583

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: Standard errors in parentheses. Weighted least squares are used to correct for heteroskedasticity. *** p< 0.01, **p<0.05, *p<0.1. The time dimension ranges from 1999 to 2017. Productivity and value added per capita are in logarithms.

Table 10.A4 The effect of governance on SOE performance

	Return on Equity	Return on Asset	Operating Profit per Sales	Labor Costs per Op. Revenue	Productivity: Sales per Employee	Value Added per Empl.
	[1]	[2]	[3]	[4]	[5]	[6]
GDP growth	0.0370	−0.0132	0.3362**	−0.7255***	−0.0948	−3.3954***
	(0.0425)	(0.0198)	(0.1457)	(0.0959)	(0.4612)	(0.4560)
Share of oil exports as a share of total exports	−0.0590***	−0.0227***	−0.1463***	0.4182***	−1.8946***	−1.1053***
	(0.0140)	(0.0066)	(0.0481)	(0.0353)	(0.1518)	(0.1732)
Liquidity: current ratio	0.0010***	0.0006***	0.0021***	−0.0004	0.0091***	0.0062***
	(0.0002)	(0.0001)	(0.0005)	(0.0004)	(0.0016)	(0.0020)
Solvency (Shareholders funds/Total assets)	−0.0127***	0.0187***	−0.0200*	0.0724***	−0.6654***	−0.1401***
	(0.0032)	(0.0015)	(0.0110)	(0.0076)	(0.0342)	(0.0349)
Total employment	0.0032***	0.0024***	0.0155***			
	(0.0004)	(0.0002)	(0.0014)			
Sales	0.0314***	0.0165***	0.0609***			−0.0095
	(0.0008)	(0.0004)	(0.0027)			(0.0111)
Total assets					0.1797***	0.1668***
					(0.0034)	(0.0046)
GDP per capita	−0.0055**	−0.0020*	0.0036	−0.0133**	0.5433***	0.9859***
	(0.0025)	(0.0012)	(0.0085)	(0.0061)	(0.0268)	(0.0261)
Transition economies (dummy)	−0.0251***	−0.0134***	−0.0680***	0.0348***	−0.7177***	−0.3241***
	(0.0027)	(0.0013)	(0.0094)	(0.0065)	(0.0300)	(0.0323)

Table 10.A4 (Continued)

	Return on Equity	Return on Asset	Operating Profit per Sales	Labor Costs per Op. Revenue	Productivity: Sales per Employee	Value Added per Empl.
	[1]	[2]	[3]	[4]	[5]	[6]
Ease of starting a business	0.0879***	0.0376***	0.1783***	−0.1069***	0.6320***	0.2696***
	(0.0073)	(0.0034)	(0.0252)	(0.0177)	(0.0802)	(0.0749)
Control of Corruption, Estimate	0.0147***	0.0086***	0.0201***	−0.0341***	0.3677***	0.2848***
	(0.0018)	(0.0009)	(0.0063)	(0.0041)	(0.0201)	(0.0157)
Industry fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.3073***	−0.1520***	−0.9056***	0.8987***	−10.7956***	−14.3814***
	(0.0419)	(0.0196)	(0.1430)	(0.0925)	(0.4509)	(0.3935)
Observations	87,860	86,316	86,078	69,215	87,360	32,696
R^2	0.2370	0.2675	0.1153	0.3376	0.8229	0.8820
Number of firms	12 729	12 670	12 525	10 888	12 638	5,858

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: Standard errors in parentheses. Weighted least squares are used to correct for heteroskedasticity. *** p< 0.01, **p<0.05, *p<0.1. The time dimension ranges from 1999 to 2017. Productivity and value added per capita are in logarithms.

Table 10.A5 The nexus between governance and ownership on firms' performance

	Return on Equity	Return on Asset	Operating Profit per Sales	Labor Costs per Op. Revenue	Productivity: Sales per Employee	Value Added per Empl.
	[1]	[2]	[3]	[4]	[5]	[6]
GDP growth	0.6067***	0.2528***	0.3773***	−0.7871***	3.7098***	-1.2921***
	(0.0067)	(0.0028)	(0.0056)	(0.0145)	(0.0472)	(0.0603)
Share of oil exports as a share of total exports	−0.0365***	−0.0163***	0.0153***	0.1888***	−2.0579***	-1.1746***
	(0.0027)	(0.0012)	(0.0023)	(0.0049)	(0.0197)	(0.0209)
Liquidity: current ratio	0.0006***	0.0000	0.0012***	−0.0016***	−0.0081***	-0.0006***
	(0.000)	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0002)
Solvency (Shareholders funds/Total assets)	−0.0292***	0.0800***	0.0303***	0.0913***	−0.2275***	0.2824***
	(0.0005)	(0.0002)	(0.0005)	(0.0009)	(0.0040)	(0.0034)
Total employment	0.0019***	0.0005***	0.0019***		0.2609***	0.1852***
	(0.0000)	(0.0000)	(0.0001)		(0.0005)	(0.0005)
Sales	0.0263***	0.0116***	−0.0084***			
	(0.0001)	(0.0001)	(0.0001)			
Total assets						
GDP per capita	−0.0099***	−0.0067***	−0.0005	0.0343***	0.7370***	1.1061***
	(0.000)	(0.0002)	(0.0003)	(0.0007)	(0.0027)	(0.0029)
Ease of starting a business	0.1785***	0.0847***	0.1003***	−0.1121***	0.4591***	0.0331***
	0.0009	(0.0004)	(0.0008)	(0.0016)	(0.0071)	(0.0060)

Table 10.A5 (Continued)

	Return on Equity	Return on Asset	Operating Profit per Sales	Labor Costs per Op. Revenue	Productivity: Sales per Employee	Value Added per Empl.
	[1]	[2]	[3]	[4]	[5]	[6]
Control of Corruption, Estimate	0.0169***	0.0180***	0.0470***	−0.0834***	0.4509***	0.3016***
	(0.000)	(0.0003)	(0.0007)	(0.0014)	(0.0060)	(0.0062)
Ownership = 1 (POEs)	0.0584***	0.0372***	0.1019***	−0.1520***	0.8261***	0.3741***
	(0.0009)	(0.0004)	(0.0008)	(0.0015)	(0.0066)	(0.0076)
Ownership = 1 (POEs) x Governance	−0.0188***	−0.0167***	−0.0531***	0.1194***	−0.2630***	-0.0973***
	(0.000)	(0.0003)	(0.0007)	(0.0013)	(0.0057)	(0.0060)
Industry fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.6755***	−0.3551***	−0.4770***	0.4763***	−12.4869***	-14.8943***
	(0.005)	(0.0024)	(0.0047)	(0.0089)	(0.0400)	(0.0372)
Observations	41 05 002	4 028 716	4 012 733	3 553 519	4,064,628	2,456,261
R^2	0.1205	0.2592	0.0869	0.1039	0.6060	0.6178
Number of firms	858 898.0	852 622	846 953	717 396	855,377	498,605

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: Standard errors in parentheses. Weighted least squares are used to correct for heteroskedasticity. *** p< 0.01, **p<0.05, *p<0.1. The time dimension ranges from 1999 to 2017. Productivity and value added per capita are in logarithms

NOTES

1. The authors would like to thank various IMF colleagues and the editors of this handbook for their comments. The views expressed in this chapter are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.
2. International Monetary Fund.
3. International Monetary Fund.
4. Although there is no commonly accepted definition of an SOE, there are some shared elements as the entity: has its own, separate legal personality; is at least partially controlled by a government unit; and engages predominantly in commercial or economic activities. Assessing government control of an entity involves judgment. A government may exercise significant influence over corporate decisions even when it owns a small number of shares. Our focus will be mainly in companies where the government owns a majority of its equity.
5. In this chapter, the focus is on non-financial SOEs. However, state-owned (public) banks are also an important part of modern economies.
6. This is particularly common and recurrent in many developing countries. For example, see Krueger (1990).
7. See IMF (2020), Musacchio and Pineda (2019), OECD (2018a), Transparency International (2018) and World Bank (2004).
8. Petri and Taube (2003) estimate quasi-fiscal activities in the energy sector at 26.7 percent of GDP in Azerbaijan in 1999 and 6.5 percent in Ukraine in 2000. Trimble et al. (2016) estimate that quasi-fiscal deficits in Sub-Saharan Africa's electricity sector are about 1.6 percent of GDP with 10 countries where it exceeds 2 percent of GDP.
9. Some of these issues also appear with private providers as governments may still constrain their ability to charge prices or collect fees. Privatization or allowing private operators also does not always deliver improvements in efficiency. In advanced economies, evidence is mixed on whether reforms delivered the anticipated efficiency gains (Gathon and Pestiau, 1995).
10. For instance, in Chile, the 2007 reform of the informal bus transportation system in the capital has been perceived as not successful. The intention was to reduce congestion, pollution and accidents through additional bus lines and fare integration with the metro (Gomez-Lobo, 2012). But the reform also required large investments that could not be covered by tariffs. The financial viability of the SOE operating the metro deteriorated rapidly, resulting in large direct subsidies from the government.
11. Some recent studies have looked at groups of European countries, including EC (2016).
12. About 15 000 are majority state owned, and 4000 are minority state owned. Due to data constraints the analysis relies mainly on firms from advanced and emerging market economies. The sample includes sectors where non-financial SOEs could potentially compete with private firms: agriculture, electricity and gas, water and sewerage, mining (including oil) and quarrying, manufacturing, communication, and construction. Adding wholesale and retail trade does not change the main results.
13. The analysis is based on financial data. For example, labor productivity is proxied by sales per employee, which does not necessarily only reflect differences in technical efficiency.
14. This result is confirmed when controlling for different factors (Table A.2–A.5 in Appendix 2). It is also consistent with Dewenter and Malatesta (2001) and more recent studies for China, Russia, and other countries in the Central, Eastern and Southeastern European region (Abramov et al., 2017; Lardy, 2019; Richmond et al., 2019). However, for Chinese SOEs, the evidence on labor productivity is mixed. Hsieh and Song (2015) find that labor productivity was similar between Chinese SOEs and private firms. However, capital productivity was much lower for SOEs.
15. The systemic difference in performance between SOEs and private firms is confirmed through several robustness checks (see the Appendix). For example, we use a two-step approach (in addition to the pooled regression) that allows us to identify both time-varying (including lagged variables) and time-invariant determinants of financial performance. The results are qualitatively and quantitatively close.
16. Goldeng et al. (2008) found similar results using Norwegian data.
17. Table A.2. in Appendix 2 contains only SOEs. The estimated coefficient for the degree of state ownership has the expected sign and is statistically significant. That is, the higher the degree of state ownership, the lower are the profits (return on equity, return on assets or profit margins), the lower is labor productivity (sales per employee) and value added per employee and the higher is labor cost per operating revenue.
18. The analysis compares SOEs with government ownership below 50 percent, above 50, and private firms (no government ownership). To this end, we created dummies for each category of firm and used SOEs with government ownership between 50 percent and 100 percent as the reference group (the baseline).
19. We use the control of corruption (CC) indicator from the Worldwide Governance Indicators (WGI). The CC is mainly based on surveys of perception of corruption (see Kaufmann et al., 2007, 2010), and available since 1996. This indicator uses information from a variety of household and business surveys. However, caution is needed in interpreting scores for any individual country given measurement error because the quality of underlying data varies across countries and data sources. Hamilton and Hammer (2018) shown that the Corruption Perception Index (from Transparency International) and the Control of Corruption are—despite some limitations—the most valid measures of the magnitude of overall corruption in many countries. These two indicators are strongly correlated.

20. Fearnley and Aarhaug (2019), using as an example the urban transportation in Norway, illustrate how complex it can be to assess which groups benefit the most from government subsidies and how to ensure different policy objectives are met (redistribution and green goals).
21. These could include indicators of (i) *profitability* (for example return on equity or assets), which provides an indication of whether the assets of the government are being well used and likely future flows to the government; (ii) *leverage* and *liquidity* to gauge the level of indebtedness of the SOEs and risk of financial distress and need for government support; (ii) *revenues and costs per worker* to help assess the efficiency of the SOE (see Baum et al., 2020).
22. India, Paraguay, the Philippines and Sweden publish reports on the aggregate performance of the SOE sector. Brazil, Ghana, India, Korea and Sweden also provide information at the individual SOE level. Note that in general these reports do not include subnational SOEs.
23. Based on IMF's Fiscal Transparency Evaluations since 2014, around 90 percent of the countries evaluated did not publish comprehensive information on the public sector.
24. The integration of SOEs in public accounts require harmonized reporting frameworks to be developed across various public entities to improve public financial management capacity.
25. The IMF's *Fiscal Transparency Handbook* recommends the disclosure of quasi-fiscal activities, including the rationale for undertaking them and the mechanisms used to compensate SOEs.
26. This indicator measures the quality of public services, civil service, policy formulation, policy implementation and credibility of the government's commitment to raise these qualities or to keep them high. Note also that the control of corruption is highly correlated with other measures of corruption, like the one from Transparency International.
27. Kalemli-Ozcan and others (2015) computed coverage ratios by dividing the value of total output produced in ORBIS (operating revenue) by the value of total output produced in a given country for a given year for European countries. For example, the average coverage of the Netherlands is low (about 26 percent) while the average coverage of Estonia is excellent (about 91 percent).
28. This estimator has been used by other authors to deal with time-invariant regressors. See Baum et al. (2019) for more technical details.
29. The results are available upon request.

REFERENCES

Abramov, A., A. Radygin, and M. Chernova. 2017. "State-owned enterprises in the Russian market: Ownership structure and their role in the economy," *Russian Journal of Economics*, 3 (1); 1–23.

Allen, R. and M. Alves. 2016. "How to Improve the Financial Oversight of Public Corporations," How-to Note, IMF Fiscal Affairs Department, International Monetary Fund, Washington, DC.

Allen, R. and S. Vani. 2013. "Financial Management and Oversight of State-Owned Enterprises," In *The International Handbook of Public Financial Management*, edited by R. Allen, R. Hemming, and B. H. Potter. Houndmills, UK: Palgrave Macmillan Limited.

Bai, C. E., J. Mao, and Q. Zhang. 2014. "Measuring market concentration in China: The problem with using censored data and its rectification," *China Economic Review*, 30 (September): 432–447.

Baum, A., C. Hackney, P. Medas, and M. Sy. 2019. "Governance and SOEs: How Costly Is Corruption?" IMF Working Paper No 19–253, International Monetary Fund, Washington, DC.

Baum, A., P. Medas, A. Soler, and M. Sy. 2020. "Managing Fiscal risks from State-Owned Enterprises," IMF Working Paper No 20–213, International Monetary Fund, Washington, DC.

Baum, A., P. Medas, A. Soler, and M. Sy. 2021. "How To Assess Fiscal risks from State-Owned Enterprises: Benchmarking and Stress Testing," Fiscal Affairs Department, *How To Notes* No 21/09, International Monetary Fund, Washington, DC.

Brollo, F., T. Nannicini, R. Perotti, and G. Tabellini. 2013. "The Political Resource Curse," *American Economic Review*, 103 (5): 1759–96.

Cuervo-Cazurra, A., A. Inkpen, A. Musacchio, and K. Ramaswamy. 2014. "Governments as Owners: State-Owned Multinational Companies," *Journal of International Business Studies* 45: 919–42.

Dewenter, K. and P. Malatesta. 2001. "State-owned and privately-owned firms: An empirical analysis of profitability, leverage, and labor intensity," *American Economic Review*, 91 (1).

European Commission (EC). 2016. "State-Owned Enterprises in the EU: Lessons Learned and Ways Forward in a Post-Crisis Context," European Commission, Institutional Paper 031–July 2016.

Fearnley, N. and J. Aarhaug. 2019. "Subsidising urban and sub-urban transport – distributional impacts," *European Transport Research Review*, 11: 49.

Gathon, H.J. and P. Pestiau. 1995. "Decomposing efficiency into its regulatory components: The case of European Union," *European Journal of Operational Research*, 80 (3): 500−507.

Gleave, S. D. 2016. "Study on the Prices and Quality of Rail Passenger Services," European Commission.
Goldeng E., L. Grünfeld, and G. Benito. 2008. "The Performance Differential Between Private and State-Owned Enterprises: The Roles of Ownership, Management and Market Structure," *Journal of Management Studies*, 45 (7): 1244–73.
Gómez-Lobo, A. 2012. "The ups and downs of a public transport reform: the case of Transantiago," Working Papers wp354, University of Chile, Department of Economics.
Hsiao, C. H. 2003. Analysis of Panel Data, UK: Cambridge University Press.
Hsieh, C.-T. and M. Song. 2015. "Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China," *Brookings Papers on Economic Activity*, Spring 2015, pp. 295–346.
International Monetary Fund (IMF). 2017. "People's Republic of China Article IV Staff Report 2017," IMF Country Report 17/247, Washington, DC.
———. 2019. "Curbing Corruption," *Fiscal Monitor*, April 2019, International Monetary Fund, Washington, DC.
———. 2020. "State-Owned Enterprises: The Other Government," *Fiscal Monitor*, May 2020, International Monetary Fund, Washington, DC.
Kaufmann, D., A. Kraay, and M. Mastruzzi. 2007. "Worldwide Governance Indicators Project: Answering the Critics," *World Bank Policy Research Working Paper* No. 4149.
——. 2010. "The Worldwide Governance Indicators: Methodology and Analytical Issues," *World Bank Policy Research Working Paper* No. 5430.
Kole, Stacey and J. Harold Mulherin. 1997. "The government as a shareholder: A case from the United States," *The Journal of Law and Economics*, 40 (1): 1–22.
Kornai, J. 1979. "Resource-constrained versus demand-constrained systems," *Econometrica*, 47 (4): 801–19.
Krueger, A. O. 1990. "Government failures in development," *Journal of Economic Perspectives*, 4 (3): 151–74.
Lardy, N. 2019. "The State Strikes Back: The End of Economic Reform in China?" Peterson Institute for International Economics, Washington, DC.
Maliszewski, W., S. Arslanalp, J. Caparusso, J. Garrido, S. Guo, J. S. Kang, W. R. Lam, D. Law, W. Liao, N. Rendak, P. Wingender, J. Yu, and L. Zhang. 2016. "Resolving China's Corporate Debt Problem," IMF Working Paper 16/203, International Monetary Fund, Washington, DC.
Megginson, M. and J. Netter. 2001. "From State to Market: A Survey of Empirical Studies on Privatization," *Journal of Economic Literature* 39 (2): 321–89.
Menard, C. and A. Peeroo. 2011. "Liberalization in the Water Sector: Three Leading Models," In *International Handbook of Network Industries: The Liberalization of Infrastructure*, edited by Matthias Finger and Rolf W Künneke, 310–327, Cheltenham, UK, and Northampton, MA: Edward Elgar.
Menozzi, Urtiga, and Vannoni. (2012). "Board composition, political connections, and performance in state-owned enterprises," *Industrial and Corporate Change*, 21 (3): 671–98.
Musacchio, A. and Emilio I. Pineda. 2019. "*Fixing State-Owned Enterprises in Latin America: Old Problems, New Solutions*," Washington, DC: Inter-American Development Bank.
Natural Resource Governance Institute (NRGI). 2019. *National Oil Company Database*. New York, NY: Natural Resource Governance Institute.
Organisation for Economic Co-operation and Development (OECD). 2014. "Foreign Bribery Report: An Analysis of the Crime of Bribery of Foreign Public Officials," Paris: OECD Publishing.
———. 2015. "*OECD Guidelines on Corporate Governance of State-Owned Enterprises, 2015 Edition*," Paris: OECD Publishing.
———. 2018a. "*State-Owned Enterprises and Corruption: What Are the Risks and What Can Be Done?*" In Paris: OECD Publishing.
———. 2018b. "*Ownership and Governance of State-Owned Enterprises: A Compendium of National Practices*," Paris: OECD Publishing.
———. 2019a. *"Measuring Distortions in International Markets: The Aluminium Value Chain,"* OECD Trade Policy Papers, No. 218, OECD Publishing.
———. 2019b. "*Measuring Distortions in International Markets: The Semiconductor Value Chain*," OECD Trade Policy Papers, No. 234, OECD Publishing.
Pérard, E. 2009. "Water supply: Public or private?" *Policy and Society*, 27 (3): 193–219.
Petri, M. and G. Tauber. 2003. "Fiscal policy beyond the budget: Quasi-fiscal activities in the energy sectors of the former Soviet Union," *Emerging Markets Finance and Trade*, 39 (1): 24–42.
Public-Private Infrastructure Advisory Facility (PPIAF). 2016. "Improving Private Sector Participation in Vietnam," *Impact Assessment*, October 2016.
Restuccia, D. and R. Rogerson. 2019. "The causes and costs of misallocation," *Journal of Economic Perspectives*, 31 (3): 151–74.
Richmond, C. J., D. Benedek, E. Cabezon, B. Cegar, P. Dohlman, M. Hassine, B. Jajko, P. Kopyrski, M. Markevych, J. A. Miniane, F. J. Parodi, G. Pula, J. Roaf, M. Kyu Song, M. Sviderskaya, R. Turk, and S. Weber. 2019. "Reassessing the Role of State-Owned Enterprises in Central, Eastern, and Southeastern Europe," *European Departmental Paper Series* 19/11, International Monetary Fund, Washington, DC.

Shirley, M. 1995. "Bureaucrats in Business: The Economics and Politics of Government Ownership," Policy Research Report 15037, World Bank, Washington, DC.
Shirley, M. 1998. "Why performance contracts for state-owned enterprises haven't worked," *Public Policy for the Private Sector*, 150 (August 1998).
Simpson, S. and H. N. Nyante. 2015. "Evaluating performance contracts as a tool for evaluating state-owned enterprises in two developing countries: A look at the performance dimensions," *Public Policy and Administration Research*, 5 (9): 27–35.
Song, Z., K. Storesleten, and F. Zilibotti. 2011. "Growing like China," *American Economic Review*, 101 (1): 202–41. Transparency International. 2018. "State-Owned Enterprises: Beacons of Integrity?".
Trimble, C., M. Kojima., A. Perez, and F. Mohammadzadeh. 2016. "*Financial Viability of Electricity Sectors in Sub-Saharan Africa: Quasi-Fiscal Deficits and Hidden Costs*." Policy Research Working Paper No. 7788. World Bank, Washington, DC.
U.S. Department of Transportation. 2019. "National Transit Summaries and Trends 2018."
Vining, A. and A. Boardman. 1992. "Ownership versus competition: Efficiency in public enterprise," *Public Choice*, 73: 205–239.
World Bank. 2004. "*Reforming Infrastructure: Privatization, Regulation, and Competition*," Washington, DC: World Bank.
———. 2017. "*Who Sponsors Infrastructure Projects? Disentangling Public and Private Contributions*," Washington, DC: World Bank.
———. 2019. "*Rethinking Power Sector Reform in the Developing World*," Washington, DC: World Bank.
World Health Organization (WHO) and the United Nations Children's Fund (UNICEF). 2017. "Progress on Drinking Water, Sanitation and Hygiene: 2017 Updated and SDG Baselines," Geneva.

PART III

PUBLIC SECTOR EFFICIENCY IN THE EDUCATION, HEALTH, AND SOCIAL PROTECTION SECTORS

11. Public spending efficiency in compulsory education

Douglas Sutherland[1]

1. WHAT IS EDUCATION SPENDING EFFICIENCY?

With government debt levels having risen as a legacy of the COVID-19 shock coupled with a secular slowing of trend growth in many (OECD) countries and ageing-related spending continuing to mount, government budgets face considerable pressure. Education accounts for a sizeable proportion of government spending, such that improving spending efficiency is an important objective (Agasisti, 2014). The link between investment in education and subsequent growth argues for maintaining spending while the pressures on public finance focus attention on minimising inefficiency without undermining educational outcomes and, as such, realising savings where possible. As a result, spending efficiency in compulsory education requires action that balances the costs of provision with the wider benefits arising from having a more educated population.

The public sector typically plays an important role in financing education in primary and secondary schools (OECD, 2017). Raising educational attainment − through the effects on the skills of the population − strengthens labour market skills, boosting productivity and economic wellbeing in the longer term (Hanushek & Woessmann, 2020). Using resources efficiently promises to pay dividends. Analysis assessing potential improvements from raising educational attainment to output growth suggest sizeable impacts (Canton et al, 2018). The benefits have encouraged governments to play an active role in primary and secondary education (hereafter, compulsory education). Raising the skills of the population in the coming decades will be important in harnessing the increasing digitalisation of economies and meeting the demands of shifting patterns of employment. Facing the headwinds created by ageing populations, compulsory education will need to remain flexible in providing the skills needed for the future.

However, measurable educational outcomes are not the sole metric to assess educational systems, particularly by parents and politicians. For example, schools are important beyond attainment of traditional academic skills, including through influencing pupil behaviour and their social attitudes, civic engagement and equity or equality of opportunity. In this light, assessing measurable efficiency for policy purposes requires an appreciation of other objectives. Furthermore, the education production function is difficult to define. Educational systems organise themselves differently making direct comparison complicated. For example, school-level analysis will take into account the resources determined by the school, such as teacher numbers, but may fail to account for centralised purchases of textbooks and other educational inputs. The relative role of the school and educational ministries in undertaking pedagogical work to develop curricula and teaching methods will also complicate comparison of schools across countries. Furthermore, schools in OECD countries may offer ancillary services – such as meals, transport, sport and housing – to support students and their families. As such, simple spending comparisons may miss differences in the functions of the school.

The operation of schools involves a host of principal-agent relationships between different stakeholders that influence school priorities and resource utilisation (for example central and local government; local government and schools; parents and teachers). The decentralisation of resources and the autonomy of resource use can mean that local preferences are better met, but spending may diverge from achieving national priorities. As a result, countries often establish monitoring and evaluation systems to ensure that resources are being used as intended (Fakharzadeh, 2016). That said, attempts to wring out efficiency gains can create tensions between pedagogical leadership and resource management responsibilities. As such, paying greater attention to one output can lead to less desirable outcomes along the lines explored in public sector multitask-principal-agent settings (Dixit, 2002). For example, attempts to constrain spending pressures emanating from teachers' salaries may prove counterproductive by dissuading highly qualified candidates from entering − and good teachers remaining in − teaching (OECD, 2020).

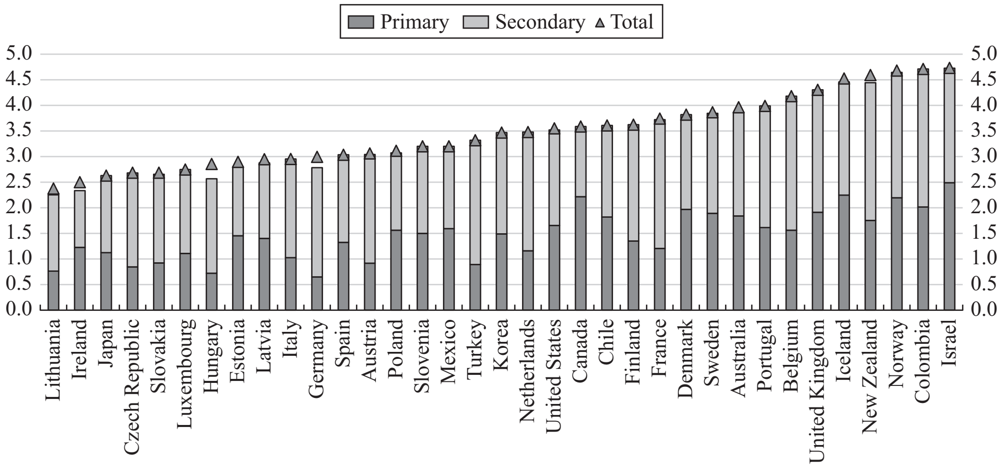
2. IMPORTANCE OF COMPULSORY EDUCATION IN PUBLIC FINANCES

Spending on primary and secondary education mainly flows to compulsory education. In OECD countries, such spending represents on average around 3½ per cent of GDP, of which the overwhelming majority comes from the public purse (over 90 per cent). However, in some countries the private sector can be an important source of funding (reaching almost one-third of spending in Turkey, for example). There is considerable variation across countries with spending in primary and secondary education equivalent to 2.3 per cent of GDP in Lithuania and rising to 4.7 per cent of GDP in Costa Rica (Figure 11.1). The use of GDP as a denominator can be misleading when there are sizeable differences with gross national income, as is the case for Ireland and Luxembourg. The amount spent on compulsory education has been squeezed over time. Around major turning points in economic cycles, spending on education has typically declined. Spending on compulsory education in the wake of the Global Financial Crisis of 2008 was generally less affected than tertiary education. However, notwithstanding declining student rolls in a number of countries, such as Japan, spending has not kept pace with GDP growth on average.

Annual spending on primary and secondary education in OECD countries was on average around \$9000−\$10 000 per student in 2017 (OECD, 2020).[2] Spending per student tends to rise with the level of education, such that on average primary education spending is somewhat lower than at the secondary level. The variation across countries, when using purchasing power parities, was considerable running from around \$3000 in Mexico to over \$23 000 in Luxembourg (Figure 11.2). Sizeable differences can also exist within a country, particularly when local governments play an important role in financing compulsory education. For example, in the United States, over 90 per cent of final government spending is at the State level, where education spending accounts for around one third of all final expenditure.

Spending on teachers accounts for much of the difference. In general, spending on teaching staff accounts for around 95 per cent of all spending on average across the OECD. Student-teacher ratios are high in Mexico whereas teacher salaries are comparatively high in high-spending countries (Figure 11.3). School systems that integrate vocational education and training into secondary education also tend to spend more per student than their peers. Large

Total expenditure on educational institutions as a percentage of GDP, 2017

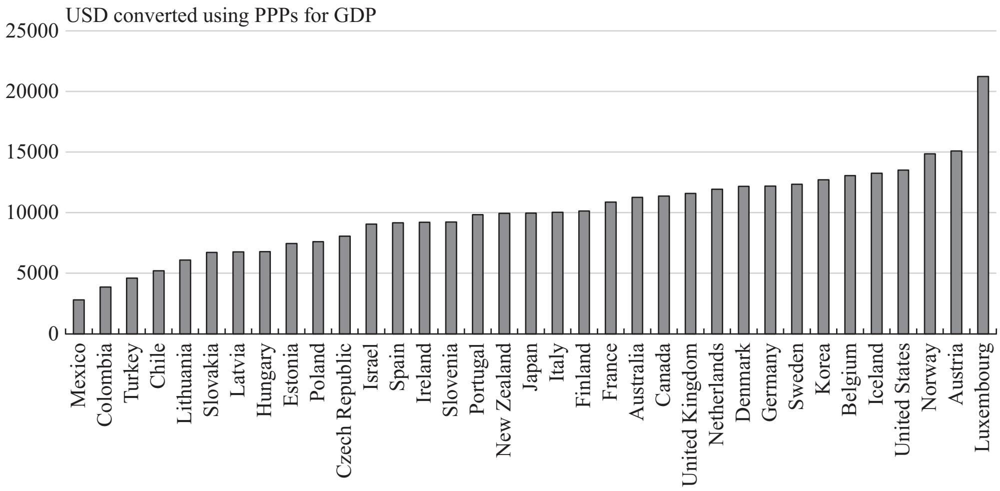


Note: Direct expenditure within primary, secondary and post-secondary non-tertiary educational institutions.

Source: OECD (2020).

Figure 11.1 Resources devoted to compulsory education can be large

Total expenditure on educational institutions per full-time equivalent student, 2017

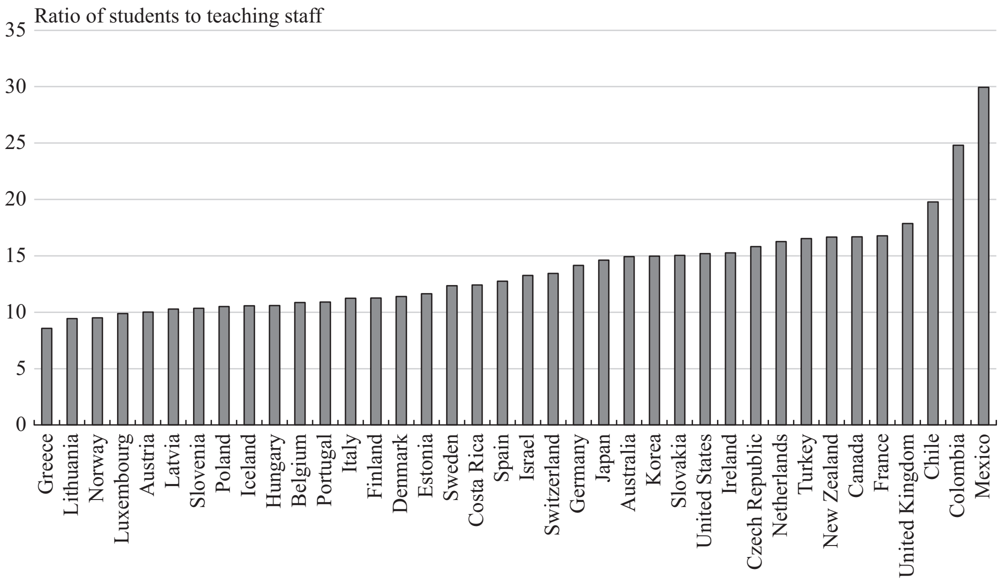


Note: Direct expenditure within primary, secondary and post-secondary non-tertiary educational institutions.

Source: OECD (2020).

Figure 11.2 Spending per pupil varies substantially across the OECD

Based on full-time equivalents



Source: OECD (2020) EAG.

Figure 11.3 Student–teacher ratios vary substantially

differences can exist across countries in ancillary services provision (such as student welfare services). For example, Denmark's schools are almost wholly dedicated to providing core education services. At the other extreme, in the Slovak Republic around one-fifth of spending is on ancillary services.

3. MEASURING PUBLIC SECTOR EFFICIENCY

Measuring public sector efficiency in education is complicated. Not only are education authorities, teachers and parents wanting a multitude of different outputs and outcomes, but, because the public sector is a major financer of education, standard metrics of success in the business sector are of limited application (Johnes, 2015). As a result, emphasis on effectiveness in meeting different objectives needs complementary analysis of whether the resources devoted to education are being used as efficiently as possible. For this, a production function with defined inputs and outputs is needed. This question lends itself to linear programming that defines frontiers and then assesses the efficiency of a 'decision making unit' relative to the empirically determined frontier.

There is a long lineage examining the efficiency of education spending at the school level. For example, Bessent and Bessent (1980) exploited results from the California Achievement Tests in 1977 to benchmark schools against similar schools, highlighting opportunities to raise educational outputs without necessarily increasing available resources. A steady stream of further studies have also examined efficiency, typically using frontier methods, mainly at the school level and within individual countries (De Witte & López-Torres, 2017; Worthington, 2010)

Measuring efficiency in the education sector is difficult due to the limitations in available data. In particular, differences in national data collection and reporting hamper international comparison. With the advent of comparable test data internationally from the Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS) (both run by the International Association for the Evaluation of Educational Achievement covering mathematics, science and literacy) and the OECD's Programme for International Student Assessment (PISA) standardised testing, benchmarking school-level efficiency across countries has developed as a separate line of inquiry. The focus of this chapter will adopt this approach, with the data used in the empirical analysis constructed from the PISA 2018 and the OECD's *Education at a Glance* databases, which are both designed so that data series are comparable across countries. However, the time-series data available at the national level are relatively limited, while the richer set of school-level data are essentially limited to cross-sections, which limits the analysis of efficiency to particular points in time. The focus is on OECD countries, although these data allow for analysis of schools in developing countries (Agasisti & Zoido, 2019)

A related strand of research, education effectiveness research has empirically assessed the influence of different influences both within the school and beyond on student outcomes, which is well summarised in Reynolds et al. (2014). Due to the contextual information required for more detailed or nuanced research, much of this work is within countries. The research on schools initially addressed whether they had a particular effect over and above the influence of a student's innate ability and socio-economic background. Subsequent work moved to identify why schools had different effects – exploring part of the black box of the education production function – and expanding output criteria to judge effectiveness beyond more easily measurable student attainment in mathematics and literacy. We will draw on this literature below.

Politicians at the national and local levels make choices affecting the provision of education. Their accountability for these choices is important in determining resource use and ultimately efficiency. However, efficiency is also influenced by a variety of factors beyond the politicians' or educationalists' direct control. For example, small schools that are typically located in more sparsely populated areas may not benefit from economies of scale available at larger, often urban, schools. There are also policies outside the compulsory education sector that may influence education outcomes. For example, early childhood interventions, such as those advocated by researchers (Heckman et al., 2013), can boost pupils' non-cognitive skills, which translates into later success in education and adult life outcomes. As such, complementarities between efficiency and equity objectives can emerge (Wößmann, 2008).

Measuring public spending efficiency is beset by a number of problems. For example, data for the variables of interest are not always available, making the reliance on proxy variables a necessity. Taking account of the many possible influences on educational outcomes can be challenging with the tools that are normally used. High dimensional production functions are not easy to estimate due to the curse of dimensionality and are often limited to cross-sections when a comparable time dimension to the data is unavailable. As such, country or school specific effects could span time-invariant influences on efficiency and thus estimates could overestimate the amount of inefficiency. Measurement error is a constraint and is particularly a problem for non-parametric methods that assume any deviation from best practice is due to inefficiency. At the school level, identifying inefficiency needs to bear in mind that schools may compete to attract better students (or push out weaker students) and that good schools may also attract better teachers. In this light, a school may appear relatively efficient but that masks differences in the quality of inputs. Value-added type models may be able to account for the potential of selection bias in students, but taking into account teacher quality, particularly in larger datasets, is challenging.

3.1 Outputs

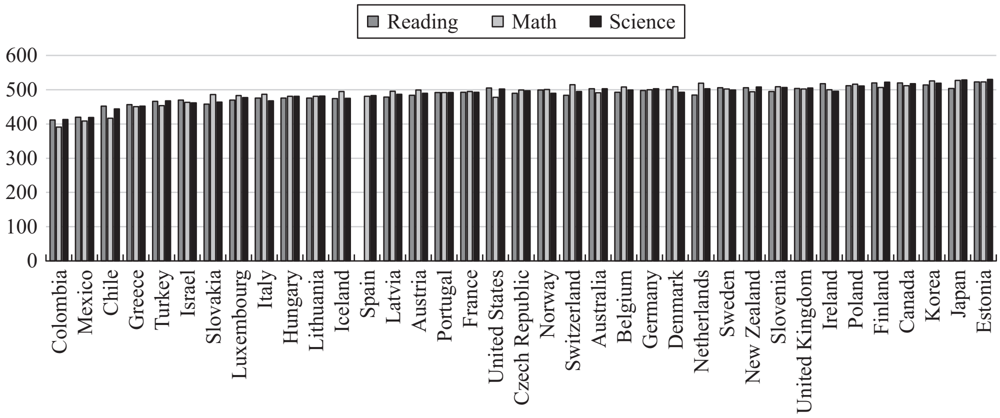
Educational output, at its most basic level, can be measured by 'quantity' indicators such as course enrolment and completion rates, study duration, the level of education attainment or even equated with the quantity of inputs. Even when test results are used, the narrow nature of academic exams has been criticised in that it may not link directly to welfare objectives or growth. Ideally, an output measure would also take into account the difference formal education makes, such as through a value-added measure, and would better capture the output. These measures are easier to construct in national testing regimes which repeatedly test pupils. An approach which takes the quality of teaching (and learning) into account would focus on outcomes, such as literacy rates at particular ages, learning achievement or longer-term earnings and occupation outcomes would give a better understanding of contributions to human capital development. However, most empirical studies of education spending efficiency rely on a narrower set of measures (Table 11.1).

For empirical work, the main output is often an indicator of the average country and school-level PISA scores. The use of PISA results for 15-year-old students which gives a measure of the cumulative output of primary and lower-secondary education goes some way to treating educational outcomes as outputs. The PISA questions are designed to evaluate real-life aptitudes and not just academic attainment (OECD, 2019). Assessment at the age of 15 is designed to capture student abilities at the end of their compulsory schooling, thus avoiding the difficulties of comparing students across countries when participation is voluntary. The country averages of the PISA scores are displayed in Figure 11.4. Students in a country that perform well in one of the areas also tend to score comparatively well in the other areas. The median

Table 11.1 Outputs considered in education efficiency studies

Quantity measures		
	Academic	Student numbers, hours of tuition, attendance rate; dropout rate
	Non-academic	Provision of non-core services
Quality measures		
	Test based	Scores, numbers passing, equity in performance
	Conditional performance	Value added measures
	Outcomes	Employment; salaries; education progression

Source: Based on De Witte and López-Torres (2017).



Note: The average country scores for the reading, math and science tests. The average of the individual tests determines the ordering.

Source: PISA 2018 database.

Figure 11.4 Average PISA scores

country score is around 493, but the dispersion across countries is quite large. The standard deviation varies from 25 to 30 points and is greater for the math scores. At the tails, the scores in Estonia are between one-quarter and one-third higher than in Colombia. This suggests at the country level a number of countries away from the tails' of the distribution are relatively similar in their scores and that the use of one score rather than composite indicators of all three scores may not make a sizeable difference.

The use of PISA results as a measure of educational output has drawbacks. National authorities may be targeting other outputs that are more difficult to take into account in the analysis. In particular, non-cognitive skills can be difficult to measure, and particularly so in an international context. As such, differences in efficiency based on a narrower set of outputs may be biased to the extent that the unmeasured outputs differ across countries and are uncorrelated with observable outputs. Secondly, in a small number of OECD countries enrolment rates are relatively low at the age the PISA tests are administered (under 60 per cent in Mexico and Turkey compared with an OECD average of 95 per cent). Cross-country comparisons could also be distorted by high truancy rates. In these cases, PISA scores may tend to overestimate the average level of human capital for the overall cohort. Related to these concerns is the fact that educational outcomes could also be judged by how many of the young continue their education to the end of the compulsory stage, which has obvious implications for human capital accumulation.

3.2 Inputs

Two main types of input determine educational outcomes, discretionary and non-discretionary (Table 11.2). The discretionary inputs include things like teachers and capital inputs, such as school buildings and IT equipment. However, particularly at the school level, getting accurate measures of the different inputs is difficult. Analysis at the school level, particularly for the

Table 11.2 Inputs considered in education efficiency studies

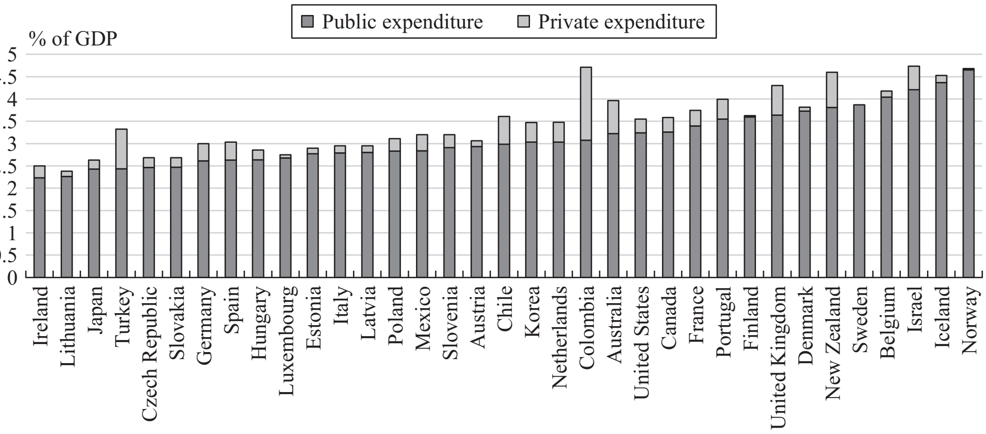
Non-discretionary		
	Individual characteristics	Age, gender, disability; native/non-native; minority status; aspirations; past test results
	Family characteristics	Social economic status; parental education; language spoken at home
	Peer effects	Characteristics of school student body
Discretionary		
	School resources	Spending; teacher numbers; student teacher ratios; teacher salaries; teacher presence; buildings; textbooks; computer and IT equipment;
	Learning environment	Attendance; school climate; dropout rates; school selectivity
	School characteristics	Size; parental involvement; location (for example urban or rural)

Source: Based on De Witte and López-Torres (2017).

United States, tends to suggest that the relationship between school resources and student outcomes is weak (Hanushek & Luque, 2003). Aggregate production functions in some ways simplify this and measures of spending as well as physical measures can facilitate spending efficiency measures more directly.

Discretionary inputs are factors under the control of the education system. They can be physical inputs, such as teacher numbers, teacher–student ratios, class sizes, instruction time, proxies for teacher quality and, to a lesser extent, other resources in schools. Discretionary inputs may also be expressed in terms of spending and the national databases report spending on primary and lower-secondary education. Spending differs somewhat more among OECD countries than the physical inputs, mainly due to disparities in unit labour costs. These disparities may reflect teaching quality and the availability of other potentially important resources available in schools, as well as labour-market factors unrelated to efficiency. This can make the use of spending problematic in the measurement of efficiency.

Financial inputs come from governments, families, firms and non-profit organisations. The government's share in compulsory education is typically high, but does vary from country to country given the different organisation of educational systems (Figure 11.5), for example, almost wholly government in Norway to a much larger share of private involvement in Columbia and Turkey and to a somewhat lesser extent in New Zealand and the United Kingdom. Public spending on compulsory education typically represents a large share of total education spending. On average spending on primary to post-secondary non-tertiary education was around 3.1 per cent of GDP for OECD countries in recent years, whereas spending on tertiary education has been around one-third of that in the OECD (around 1 per cent of GDP). The dispersion of spending across countries is large, ranging from just over 2 per cent of GDP in Ireland (reflecting the importance of multinationals in Irish national accounts) to around 4.6 per cent in Norway. Spending has been squeezed over time, dropping by over



Note: In 2017. Data on the private sector are missing for Greece.

Source: OECD (2020).

Figure 11.5 Public and private spending on education

0.2 percentage points of GDP since the global financial crisis (a somewhat smaller drop occurred in the tertiary sector as well).

Non-discretionary inputs cover environmental inputs which are not amenable to direct control by the decision-making unit, in this case either the school or the education authority. Typically, student achievement is considered to be related to innate ability and dependent on family and peer-group effects. While difficult to measure, these factors are often proxied by measures of socio-economic status, and in some cases also indicators of immigrant or language status. Typically, research suggests a student's background has a large bearing on student outcomes (Hanushek & Luque, 2003). Individual characteristics cover basic demographic factors to indicators, such as past test results, aiming to capture more accurately the value added of a school.

4. POTENTIAL INFLUENCES ON PUBLIC SPENDING EFFICIENCY

Factors influencing public spending efficiency is an area of active investigation and the analysis can be broadly grouped into factors at different levels of intervention. At the school level, beyond the provision of inputs, the policies for teaching and creating an environment in which students can learn can potentially affect the efficiency of spending.

- First, factors within the school. The make-up of the student body has been identified as correlated with efficiency, presumably through peer effects (Agasisti & Zoido, 2018). More homogenous schools appear to perform somewhat better. Schools with a higher share of female students also appear to be more efficient.
- Second, organisation of schools is also correlated with school-level efficiency. This includes whether the school is more academic or vocational orientated, the ownership and main funding source of the school (Wößmann, 2007).

- Third, school ownership is frequently considered as a determinant of differences in efficiency (Sutherland et al., 2009). For example, an approach merging propensity score matching and stochastic frontier analysis for Spanish schools using PISA data suggests that private schools that are dependent on public financing are more efficient than public schools, but this result differs in importance across regions (Crespo-Cebada et al., 2013).
- Fourth, the autonomy of schools in the education system. Given asymmetric information, greater school autonomy is seen as a possible way to boost spending efficiency. In particular, the quality of the teaching staff can be better harnessed when schools have greater autonomy in personnel management. The degree of decentralisation in this context is seen to be conducive to raising spending efficiency (Coelho, 2009).
- Fifth, the accountability of schools. Potential factors include external testing and accountability mechanisms (Wößmann, 2008). However, accountability mechanisms which have funding implications have also been found to induce strategic behaviour by schools, which may undermine the equity goals of education funding.

Elucidating the effects of different characteristics requires comparison between the public sector and private sector within a country, which may not give a truly comparable picture, or require comparison across countries, which is also complicated by the different educational and other objectives.

5. MEASURING EFFICIENCY

Most attempts to estimate education sector efficiency have used frontier-type methods, either non-parametric (such as data envelopment analysis) or parametric (such as stochastic frontier analysis) methods. Essentially, these methods differ in the assumptions needed to define a production function, which is traded off against the sensitivity of results to outliers or potential mismeasurement. The attraction of non-parametric measures arises as outputs in the public sector are amorphous and intangible in many respects. This makes it difficult to define a production function in the conventional sense, while public sector bodies producing goods that are free at the point of use means that the prices of outputs are not determined by market forces. Economic efficiency in these circumstances cannot be measured directly and approaches are needed to proxy for an efficiency frontier to facilitate benchmarking. This section looks at two main approaches for achieving this: the first is a non-parametric technique and the second stochastic frontier analysis.

5.1 Non-parametric Measures

A large share of studies examining education efficiency rely on non-parametric approaches, principally data envelopment analysis, measuring performance relative to a measure of best practice. In essence, this technique assesses performance relative to estimates of the additional output or the resource savings that are possible if the country or school increased efficiency to the level of best practice. Inputs are described both in physical terms, the relationship between input volumes (for example teachers) and outputs being a measure of technical efficiency, and in terms of the amount spent per pupil, which determines cost efficiency. A considerable attraction of non-parametric approaches to estimating efficiency is the relative simplicity of the procedure, facilitating direct measurement of efficiency for schools or more aggregate groupings. For example, non-parametric approaches do not require assumptions about the specific functional form of the educational production function, although they do assume that the production function is common to all units. In addition, the framework can easily accommodate multiple outputs, which is an attractive feature for public services where equity or other objectives may be important.

5.1.1 Data envelopment analysis

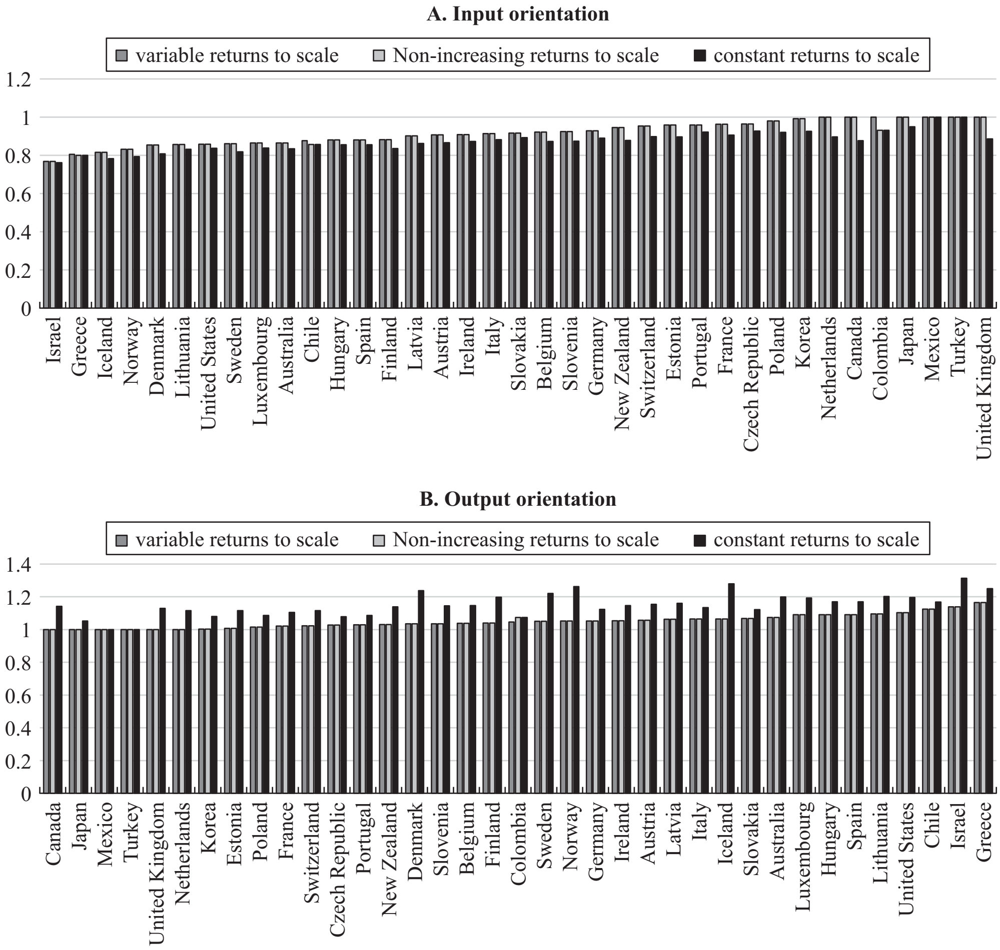
Data envelopment analysis (DEA) is one of the most common approaches to determining efficiency estimates for the public sector. This technique constructs an efficiency frontier based on the input and output data by constructing a frontier of the schools or countries that envelop the remaining observations. The frontier provides a benchmark by which the others can be judged. By assumption, the frontier determines best practice, and potential efficiency gains for specific countries or schools are measured by their position relative to the frontier or the envelope. Other assumptions include no free production, free disposability and non-infinite outputs for finite inputs.

A number of other choices are needed in determining efficiency. First, whether efficiency is measured in the input or output orientation. For example, the shortfall in terms of unachieved output is given by the ratio of a school's output to the output on the frontier for the same level of inputs. Conversely, the ratio of inputs on the frontier to the school's inputs at the same output is a measure of inefficiency in terms of potentially excess inputs. In the case of multiple inputs or outputs, the measures of efficiency are determined in a similar fashion by holding the relative proportions of either inputs or outputs constant in measuring the distance to the frontier. The efficiency frontier also depends on the assumptions about returns to scale, which can have implications for the amount of inefficiency identified:

- *Constant returns to scale (CRS)*. In a simple one input one output formulation this assumption describes the efficiency frontier as a ray from the origin through the observation(s) with the highest output/input ratio.
- *Variable returns to scale (VRS)*. This approach identifies the schools or countries that define the frontier by starting from the observations of units that use the least of each input and ending with the observations producing the highest amount of each output.
- *Non-increasing returns to scale (NIRS)*. This assumption combines the constant returns to scale assumption between the origin and the observation with the highest output/input ratio, and variable returns to scale thereafter.

To illustrate this approach we exploit the PISA 2018 database, both for OECD countries and schools within these countries, and education-at-a-glance databases to construct measures of inputs. At the country level, we take as a measure of output the scores on the different tests for math, science and reading competence. The input measures are either teacher–student ratios or different spending measures augmented by a measure of the average socio-economic status of the student's parents. At the school level, spending measures are unavailable and thus the estimation uses different measures of the tests, and as inputs the average school average socio-economic status of the student's parents and the teacher–student ratio.

The country-level results are shown in Figure 11.6 for all the assumptions about input or output variation and returns to scale and using the teacher–student ratio as the discretionary



Note: Data envelopment analysis estimates.

Source: Calculations based on PISA 2018 data.

Figure 11.6 Country-level estimates of technical efficiency

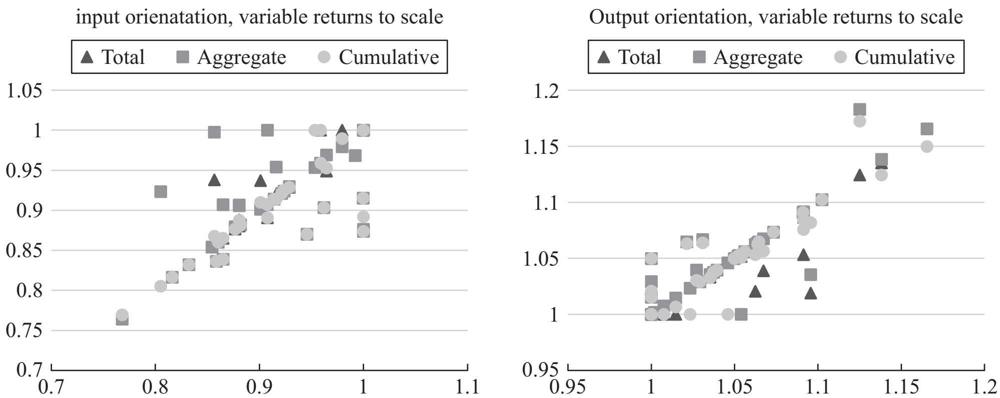
input. The results for the different assumptions are reasonably stable. The measure of efficiency when assuming constant returns to scale is much larger (around three times that of the other measures in the output orientation and 50 per cent larger in the output orientation), reflecting Mexico and Turkey determining the constant returns to scale frontier. Although PISA scores are comparatively low in these countries, the low level of inputs makes them relatively efficient. As noted above, these countries may be atypical in that school attendance is substantially lower than the OECD average. Against this background, we concentrate on the variable returns to scale case, which appears to provide a more suitable benchmark for countries on different segments of the efficiency frontier. In that case and on average for

OECD countries, the measures of efficiency suggest inputs could be reduced by around 9 per cent while maintaining output (PISA scores) constant. On the other hand, potential output gains when moving to the frontier would be relatively muted with gains of around 5 per cent holding inputs constant, reflecting the greater variation in inputs across many countries than outputs.

At the country level, efficiency estimates can replace the discretionary measure of educational inputs proxied by the ratio of teachers to students by different measures of spending. These are aggregate spending as a share of GDP, total spending per student and the cumulative spending per student between the ages of 6 and 15. The efficiency scores are reasonably correlated with the estimates based on the teacher–student ratio (Figure 11.7). The relationship is weaker for aggregate spending as a share of GDP than the other measures in the input orientation and total spending per student in the output orientation. Overall, the results suggest the variations in economic and social status are important determinants for measured efficiency, but not in all cases.

The PISA databases also provide school-level information, which is arguably closer to the decision-making unit that DEA and related measures are designed to examine. While the information is rich, there are no data on financial resources available at the school level. The sample includes nearly 9000 schools across the OECD with the sample varying from just 39 schools in Lithuania to 959 schools in Spain. Limiting the inputs and outputs to the simple technical efficiency estimates using the teacher–student ratio, the rebased measure of economic and social status and a measure of performance in the math questionnaire, the efficiency estimates show a greater degree of inefficiency. In this very simple set-up, the average school would be able to reduce input use and raise outputs by over one-third by moving to the best practice frontier. Obviously, this reflects the simplicity of the assumed production function and the

Scatter diagrams for the different spending measures compared with the technical efficiency estimates using the teacher–student ratio.

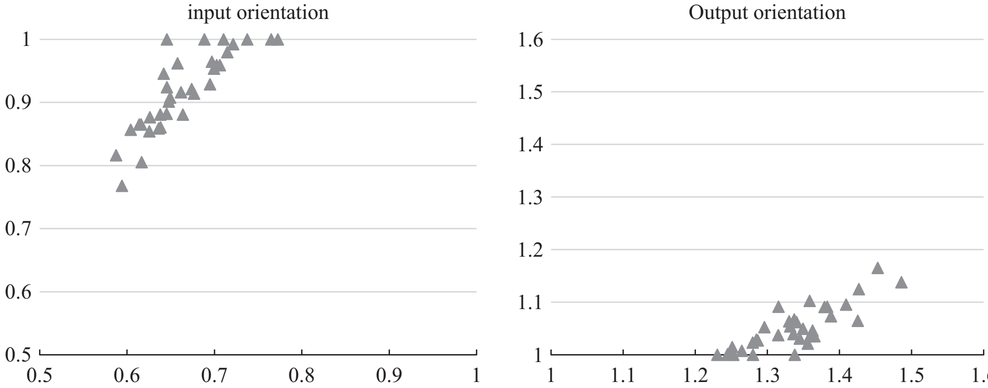


Note: Total is total spending per pupil; aggregate is spending on primary and secondary non-tertiary education; cumulative is spending on a student between the ages of 6 and 15.

Source: Calculations based on OECD PISA 2018 data.

Figure 11.7 Efficiency measures are correlated when using different spending measures

DEA estimates of efficiency using variable returns to scale



Note: The inputs are the teacher−student ratio, a measure of socio-economic status and the output is the average math test score.

Source: Calculations based on PISA 2018 data.

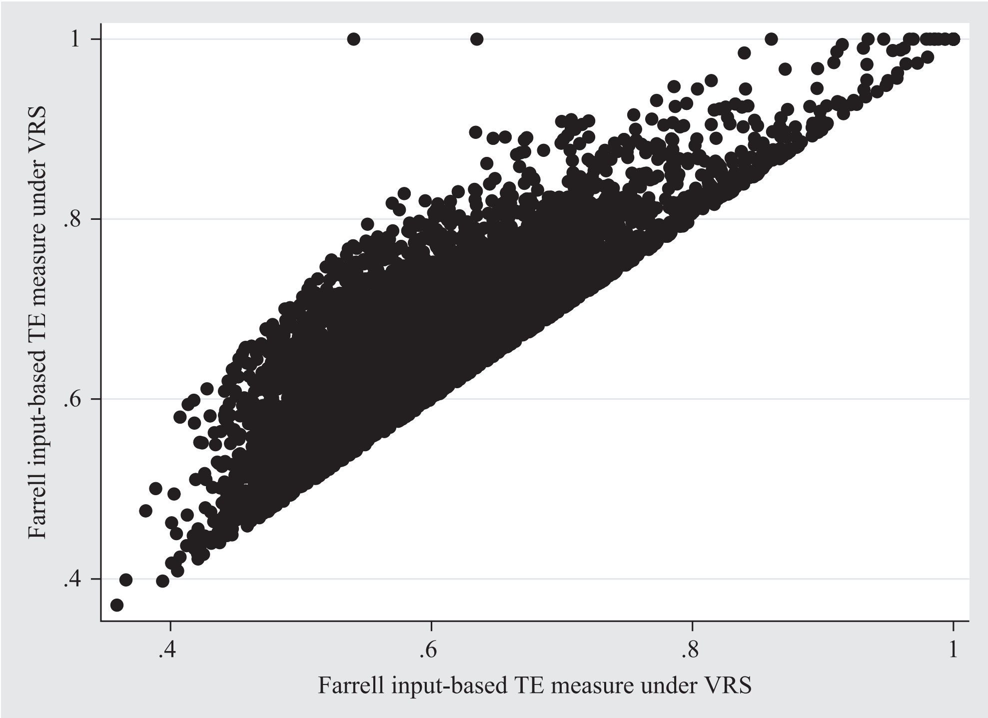
Figure 11.8 Country and median school-level efficiency estimates are similar

possibilities of extreme observations on the remaining schools. Nonetheless, the country-level estimates of efficiency are correlated with the average estimate efficiency at the school level (Figure 11.8). When an equity element is also introduced (the inverse of the interquartile range of scores on the math questionnaire) the measure of inefficiency declines. However, the measures remain strongly correlated (the correlation coefficient is 0.87) (Figure 11.9). For most countries the relationship is considerably tighter, but is weaker in Portugal, Slovakia, Spain, Estonia and Israel.

Non-parametric approaches often have a number of drawbacks that are aggravated in small samples and as the number of inputs and outputs rises. First, the technique is sensitive to measurement error and statistical noise. For example, an observation that erroneously determines a segment of the efficiency frontier will in turn affect the measures of inefficiency of all the schools or countries lying within that segment. Second, small samples − by frequently excluding best practice − can lead to a bias of under-estimating inefficiency. Third, omitted or irrelevant inputs will tend to over (or under) estimate the 'true' degree of inefficiency. In order to counter these problems, a variety of different approaches have been developed to correct for possible small-sample bias and generate confidence intervals surrounding the point estimates. Problems also arise with the curse of dimensionality. As more inputs and outputs are included in the analysis the computational demands increase exponentially, but importantly also create greater sparseness in the observations, such that estimates of the frontier become less certain.

As a reaction to the dimensionality concerns, alternative partial frontier methods offer less sensitivity to extreme observations and can ease computational requirements (Gnewuch & Wohlrabe, 2018). For example, *order M* estimators construct the frontier from a group of peers rather than the full sample. This approach allows for super-efficient schools (that is, they do not determine the frontier), which reduces efficiency estimates' sensitivity to outliers.

DEA estimates of efficiency using variable returns to scale



Note: The inputs are the teacher–student ratio, a measure of socio-economic status and the output is the average math test score and also on the vertical axis the inverse of the interquartile range of math test scores.

Source: Calculations based on PISA 2018 database.

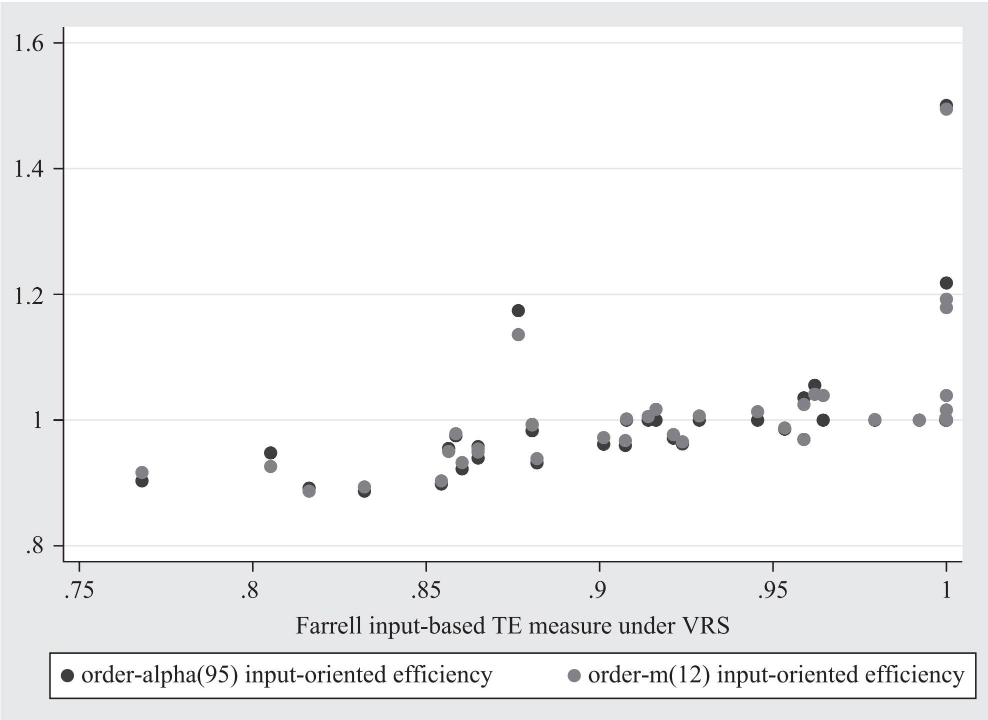
Figure 11.9 Efficiency estimates when equity is also an objective

An alternative approach, *order α*, uses a quantile-estimator approach to reduce sensitivity to outliers and permit rapid estimation. Another approach is to use statistical techniques to address the sensitivity of the estimates to where they are located on the frontier and the extent to which other observations define the frontier (Simar & Wilson, 2007).

Using these techniques on the country-level sample gives an indication of the impact on efficiency estimates (Figure 11.10). A number of countries, notably Chile, Mexico and Turkey are identified as 'super-efficient'. As a result, the efficiency frontier for other countries omits the influence of these countries. The correlation with the uncorrected estimates is apparent though weaker and the degree of identified inefficiency is correspondingly less pronounced. The least efficient countries are largely the same irrespective of the technique.

Another approach to address the robustness of non-parametric analysis is the bias-corrected approach developed by Simar and Wilson (Simar & Wilson, 2007). Bootstrapping techniques minimise the impact of unusual observations and put less emphasis on estimates when

Correlation of standard DEA measures of efficiency and alternative estimators that are less sensitive to extreme observations



Note: The inputs are the teacher–student ratio, a measure of socio-economic status and the output is the average math test score.

Source: Calculations based on OECD PISA 2018 data.

Figure 11.10 Taking into account sensitivity to extreme observations

observations in the portion of the best-practice frontier are relatively sparse, allowing a statistical correction to the potential bias in the estimate of inefficiency. This approach also permits the examination of factors associated with inefficiency as part of the analysis. To demonstrate this approach we augment the estimation with a two-stage examination using school type, school size and the homogeneity of languages spoken at home as potential factors influencing efficiency.

The results suggest that large schools, private schools and schools with more homogenous student bodies in terms of language spoken at home are also more efficient when considering input orientation efficiency (Table 11.3). In the output orientation, the coefficients remain statistically significant but now school size and language homogeneity are negatively related to the schools' estimated efficiency. The results suggest that the degree of inefficiency is somewhat larger when correcting for potential bias, although they remain strongly correlated

Table 11.3 Influences on the measure of efficiency

	Input		Output	
	coefficient	z-score	coefficient	z-score
Government-dependent private schools	0.0482	14.74	0.042514	4.18
Government-independent private schools	0.0422	14.96	0.098434	11.41
School size	0.0000	30.55	−0.0001	−24.54
Language	0.0551	8.37	−0.35972	−17.08
Constant	0.3550	51.89	1.826072	84.41
Sigma	0.0557	117.76	0.168568	114.84
N	7548		7548	

Note: The table presents the bootstrapped estimates of determinants of the degree of school-level efficiency using the Simar-Wilson approach.

Source: Calculations based on OECD PISA 2018 data.

(Figure 11.11). In general, efficiency is estimated to be higher in the output orientation and the impact of bias correction is less important. The variables included in the second stage estimates are statistically significant.

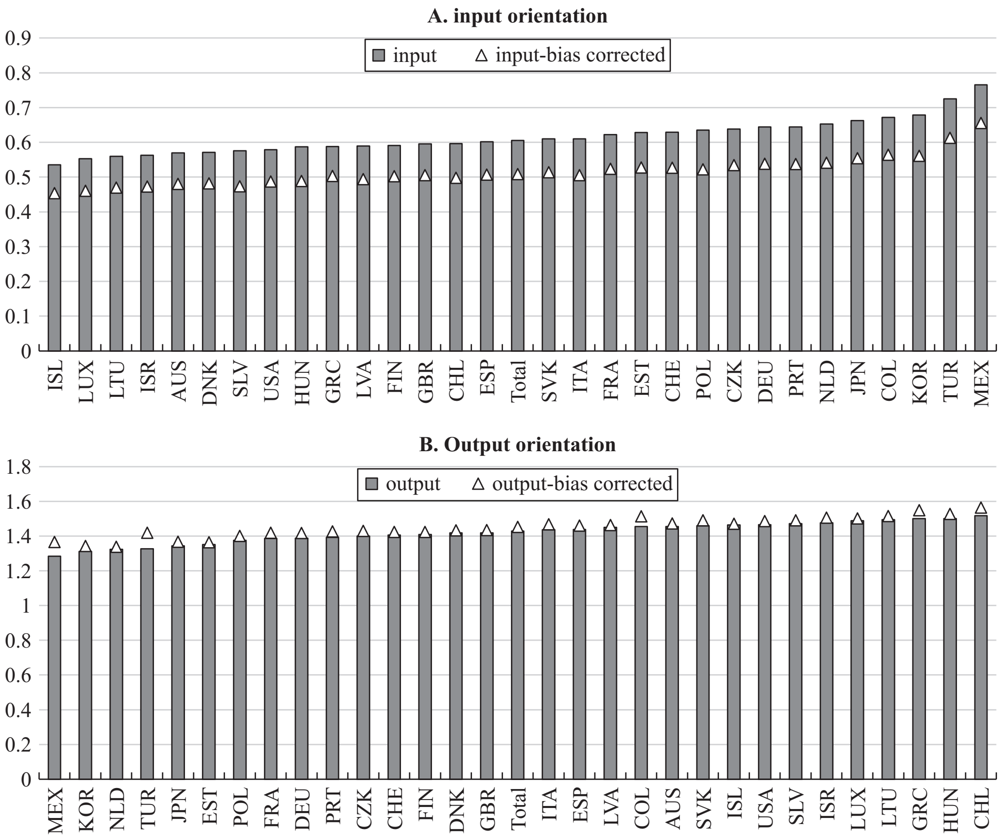
More recently, researchers have begun exploiting the growing time-series dimension in internationally comparable databases. For example, pseudo-panels constructed from PISA data have been used to examine efficiency using a Malmquist-type index within Spain, finding that private government-dependent schools are typically more efficient than public schools (Aparicio et al., 2017).

5.2 Parametric Frontier Analysis

The parametric estimation technique, stochastic frontier analysis, makes assumptions about the distribution of inefficiency decomposing the regression error term into statistical noise and a measure of inefficiency. Alternatively, fixed effects in panel models can account for country-specific inefficiencies. Stochastic frontier analysis is similar to standard regression techniques but differs by exploiting the one-sided nature of inefficiency to decompose the error term into a standard error term and an asymmetric component that measures inefficiency as shown below. The principal advantage of this approach to measuring efficiency is that it addresses statistical noise explicitly. Furthermore, and in contrast to non-parametric approaches, standard statistical tests can be used to assess variables.

Stochastic frontier analysis is less sensitive to the influence of measurement error than non-parametric approaches. In addition, the parameters allow decomposition of efficiency. However, this feature comes at the cost of specifying a functional form and making assumptions about the distribution of inefficiency. The cost of specifying a functional form can be mitigated somewhat by the choice of a flexible specification, such as a translog production function. However, estimating this production function in practice is often complicated by multicollinearity (the variables are too highly correlated with one another to allow their

Median estimates of efficiency for schools in the OECD wide sample



Note: The inputs are the teacher–student ratio, a measure of socio-economic status and the output is the average math test score.

Source: Calculations based on OECD PISA 2018 data.

Figure 11.11 School-level efficiency estimates

individual effects to be estimated with precision). In addition, the specification may not make economic sense for all observations and thus needs to be checked.

To illustrate the use of Stochastic frontier analysis, a Cobb-Douglas production function is estimated using the PISA 2018 data used in the preceding examples. In this case, a single output (a math plausible value) and several inputs (average socio-economic status of the student body, the ratio of teachers to students, the provision of computers to students, the school size and the share of students that use the language of instruction at home) are used. The estimation results suggest that academic performance is higher when the student's socio-economic status is higher, and that this correlation is very significant (Table 11.4). Test scores are also higher for larger schools, when the teacher–student ratio is higher and when more students use the

Table 11.4 Socio-economic status is an important driver of performance

Estimates from a Cobb-Douglas production function with different assumptions about the efficiency distribution

	Exponential		Half-normal	
	coefficient	z-score	coefficient	z-score
Constant	4.413	204.99	4.433	200.10
Socio-economic status	0.692	86.78	0.692	85.12
Teacher−student ratio	0.011	4.45	0.010	4.00
Computer provision	−0.011	−9.16	−0.010	−8.37
School size	0.006	3.77	0.007	4.24
Language at home	0.010	4.64	0.011	4.78
Lambda	0.84		1.68	
n	8027		8027	

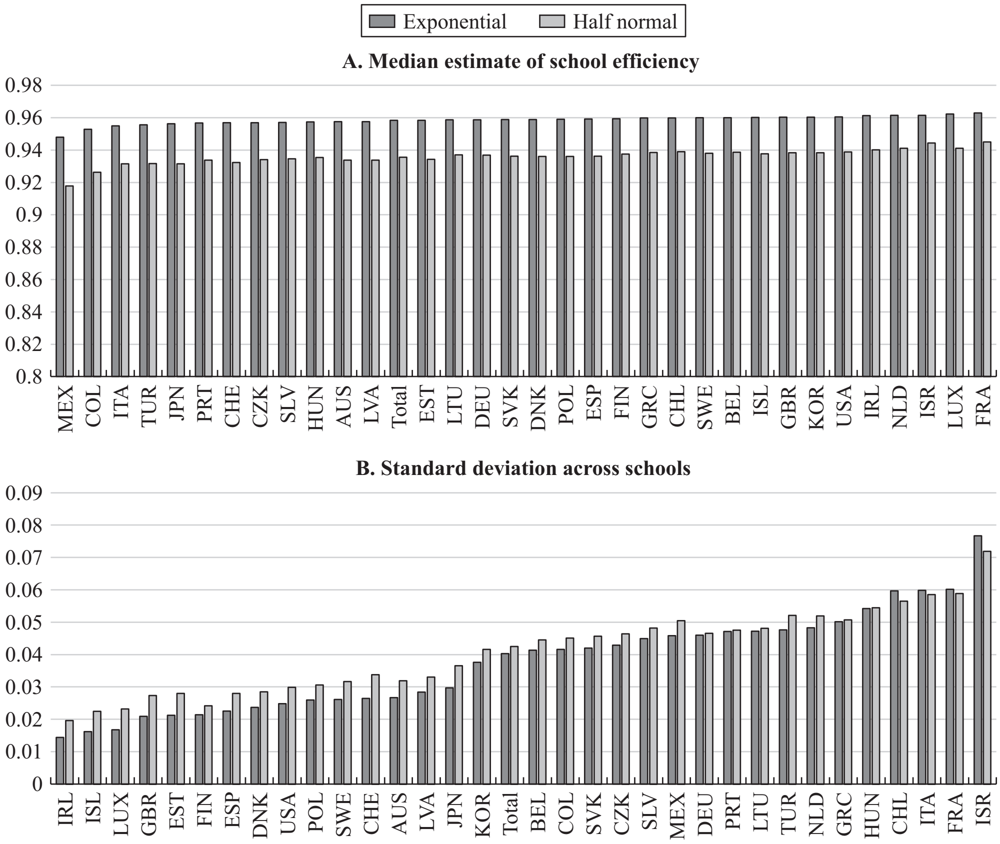
Source: Calculations based on OECD PISA 2018 data.

language of instruction at home. On the other hand, the relationship with computer availability appears to be associated with lower scores, other things being equal.

The finding of limited contribution of formal schooling resources to educational outputs is a relatively frequent finding. For example, a study of schools in Kansas found that discretionary inputs at the school (that is, those under control of school decision makers) had little impact on student achievement. On the other hand, socio-economic measures were seen to be influences on student achievement (Chakraborty, 2009). As a result, measured inefficiencies can overstate the ability of schools to improve outputs due to environmental factors, over which the school has limited influence (at least in the short term) (Henderson et al., 2017).

The estimates of the degree of inefficiency are relatively small in the estimations (Figure 11.12). Outputs could be improved by around 5−8 per cent on average. In most cases, the efficiencies of the schools are tightly clustered, particularly when the degree of inefficiency is low at the country level. For example, schools in Estonia and Japan are on average among the most efficient and the standard deviations of efficiency scores are amongst the lowest in the sample of OECD countries. In some cases, sizeable across-school heterogeneity in efficiency estimates suggests even benchmarking performance within a country may potentially offer possibilities to increase the efficiency of the education sector. For example, the estimated efficiency of the median school in Chile, Greece and Israel is relatively low. However, there is wide variation across schools in these countries. In other cases, such as France and Italy, school-level efficiency, which is already comparatively very high, could be raised even further if the poorest performing schools moved towards best practice within the country. Introducing country fixed effects into the estimation generally reduces the estimated degree of inefficiency.

Median estimates of efficiency for schools in the OECD-wide sample and standard deviation by country



Note: The results show the estimated school efficiency in a Cobb-Douglas production function using as inputs the teacher−student ratio and a measure of socio-economic status with the output being an average math test score.

Source: Calculations based on OECD PISA 2018 data.

Figure 11.12 Stochastic frontier estimates of efficiency are generally higher

Finally, the parametric approach lends itself more easily to panel data examination and thus the ability to control for time-invariant omitted variables in the estimation of the degree of efficiency. At present, national-level databases are best placed to exploit these approaches. For example, a study using a rich database of Finnish secondary schools and matriculation exam results takes into account past student performance and comes closer to a value-added-type model (Kirjavainen, 2012). The results, while confirming the relatively limited dispersion of efficiency levels across Finnish schools, also demonstrate that such efficiency measures can nonetheless be useful in identifying where performance can be improved.

6. DISCUSSION

Pressure on fiscal policy will likely make exploiting potential efficiency gains in the provision of education ever more important. A simple reading of the results presented above suggests that moving to observed best practice could save around 5–10 per cent of existing resources or boost educational outputs by a similar though smaller amount when considering national data envelopment analysis or school-level stochastic frontier analysis results. In the simple set-up used in the examples, the potential resource savings or output gains are substantially larger when using data envelopment analysis at the school level. However, while differences in efficiency can highlight where lessons from elsewhere may be helpful, improving spending efficiency is complicated as the school or education authority may not have direct control over all elements of the production function. As such the broad findings of efficiency estimates need to be complemented by more detailed analysis.

The relationship between the efficiency estimate of the median school and the country-level results suggests that the more disaggregated results remain useful. Furthermore, school-level analysis allows the assessment of more detailed policy settings which can help identify ways to improve public spending efficiency in compulsory education. Benchmarking and sharing of good practice are foundations on which to improve public spending efficiency (OECD, 2021). The empirical literature has made significant strides in adopting new techniques to minimise the potentially distorting impact of unusual schools in the sample and can be easily integrated into future analysis. As such, the approaches outlined can offer fairly robust benchmarks to help inform policy.

Ongoing work identifying how effective school systems operate (OECD, 2017) and the work of the education effectiveness literature, which has been running in parallel to public sector efficiency work, offer new insights to identifying factors driving public spending efficiency. As more comparable cross-country data become available, it will be important to examine whether this can help shed light on generalizable lessons from the detailed country-specific case study evidence. In many cases the factors are likely to boost productivity and therefore efficiency while minimising pressure on public budgets

Areas of future research include examining complementarities between investments in different levels of education. For example, the work of Heckman and others examines the importance of early childhood education on the subsequent acquisition of cognitive and non-cognitive skills, particularly for students from disadvantaged backgrounds (Heckman et al., 2013). Some work has begun to explore non-cognitive measures on school-level efficiency (Agasisti & Zoido, 2018). However, the longer-term outcomes of early childhood interventions may not be easily discernible, underlining the importance of considering wider influences of education spending when assessing efficiency results. Indeed, reforms take time to show up in both school-based data and especially in population data (employment, earnings, productivity, criminality) due to time exposure to reforms of students and the movement of cohorts into the wider population. In this light, it is important work to follow up on how education interventions play out over time. Another area of potential research is the impact of digitalisation on spending efficiency in the education sector. Education systems have reacted to the COVID-19 pandemic by making greater use of digital technologies and distance learning. For example, the Japanese government accelerated the rollout to ensure that all school students had access to computers. The lessons from the different approaches adopted across countries may be important in offering teaching environments that can overcome distance-learning shortcomings. This not only improves spending efficiency but potentially makes countries more resilient in the face of future shocks.

NOTES

1. The views expressed in this document are solely those of the author and do not necessarily represent the official views of the Organisation for Economic Co-operation and Development.
2. This measure is highly correlated with an OECD measure of cumulative spending on students between the ages of 6 and 15, which given the output measure of PISA results from 15-years-old students is an appropriate measure of spending.

REFERENCES

Agasisti, T., 2014. The Efficiency of Public Spending on Education: an empirical comparison of EU countries. *European Journal of Education*, 49(4), 543–557.
Agasisti, T. & Zoido, P., 2018. Comparing the Efficiency of Schools Through International Benchmarking: Results From an Empirical Analysis of OECD PISA 2012 Data. *Educational Researcher*, 47(6).
Agasisti, T. & Zoido, P., 2019. The efficiency of schools in developing countries, analysed through PISA 2012 data. *Socio-Economic Planning Sciences*, 68, 100711.
Aparicio, J., Crespo-Cebada, E., Pedraja-Chaparro, F. & Santín, D., 2017. Comparing school ownership performance using a pseudo-panel database: A Malmquist-type index approach. *European Journal of Operational Research*, 256(2), 533–542.
Bessent, A. M. & Bessent, E. W., 1980. Determining the Comparative Efficiency of Schools through Data Envelopment Analysis. *Educational Administration Quarterly.*
Canton, E., Thum-Thysen, A. and Voigt, P., 2018. "Economists' Musings on Human Capital Investment: How Efficient is Public Spending on Education in EU Member States?", European Economy Discussion Paper No. 081.
Coelho, M., 2009. The Effect of Organisational Structure on Education Efficiency: Public-Private Provision and Decentralisation, Institute of Local Government Studies, The University of Birmingham.
Chakraborty, K., 2009. Efficiency in Public Education - The Role of Socio-Economic Variables. *Research in Applied Economics*, 1(1), pp 1−18.
Crespo-Cebada, E., Pedraja-Chaparro, F. & Santín, D., 2013. Does School Ownership Matter? An Unbiased Efficiency Comparison for Regions of Spain. *Journal of Productivity Analysis*, 41(1), pp. 153−172.
De Witte, K. & López-Torres, L., 2017. *Efficiency in Education: A Review of Literature and A Way Forward.* s.l.Palgrave Macmillan Ltd.
Dixit, A., 2002. Incentives and Organizations in the Public Sector: An Interpretative Review. *Journal of Human Resources*, 37(4), 696–727.
Fakharzadeh, T., 2016. "Budgeting and Accounting in OECD Education Systems: A Literature Review", *OECD Education Working Papers, No. 128*, OECD Publishing, Paris.
Gnewuch, M. & Wohlrabe, K., 2018. Super-Efficiency of Education Institutions: An Application to Economics Departments. *Education Economics*.
Hanushek, E. A. & Luque, J. A., 2003. Efficiency and Equity in Schools Around the World. *Economics of Education Review*, 22, 481–502.
Hanushek, E. A. & Woessmann, L., 2020. Education, Knowledge Capital, and Economic Growth. In: *The Economics of Education: A Comprehensive Overview.* s.l.Academic Press, pp. 171–182.
Heckman, J., Pinto, R. & Savelyev, P., 2013. Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, 103(6), 2052–2086.
Henderson, D. J., Simar, L. & Wang, L., 2017. The Three Is of Public Schools: Irrelevant Inputs, Insufficient Resources and Inefficiency. *Applied Eonomics*, 49(12), 1164–1184.
Johnes, J., 2015. Operational Research in Education. *European Journal of Operation Research*, 243, 683–696.
Kirjavainen, T., 2012. Efficiency of Finnish General Upper Secondary Schools: An Application of Stochastic Frontier Analysis with Panel Data. *Education Economics*, 20(4), 343–364.
OECD, 2017. *The Funding of School Education: Connecting Resources and Learning.* s.l.OECD Publishing.
OECD, 2019. PISA 2018 Results (Volume I): What Students Know and Can Do. OECD Publishing, Paris.
OECD, 2020. *Education at a Glance 2020: OECD Indicators*, OECD Publishing, Paris.
OECD, 2021. *Fiscal Federalism 2022: Making Decentralisation Work.* OECD Publishing, Paris.
Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C. & Stringfield, S., 2014. Educational Effectiveness Research (EER): A State-of-the-Art Review. *School Effectiveness and School Improvement*, 25(2), 197–230.
Simar, L. & Wilson, P. W., 2007. Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes. *Journal of Econometrics*, 136(1), 31–64.

Sutherland, D., Price, R. & Gonand, F., 2009. Improving Public Spending Efficiency in Primary and Secondary Education. *OECD Journal: Economic Studies*, 2009, 1–30.

Worthington, A. C., 2010. "Education Economics An Empirical Survey of Frontier Efficiency Measurement Techniques in Education", *Education Economics*, 9(3), 245−268 .

Wößmann, L., 2007. International Evidence on School Competition, Autonomy, and Accountability: A Review. *Peabody Journal of Education*, 82(2–3), 473–497.

Wößmann, L., 2008. Efficiency and Equity of European Education and Training Policies. *International Tax and Public Finance*, 15, 199–230.

12. The efficiency of higher education institutions and systems

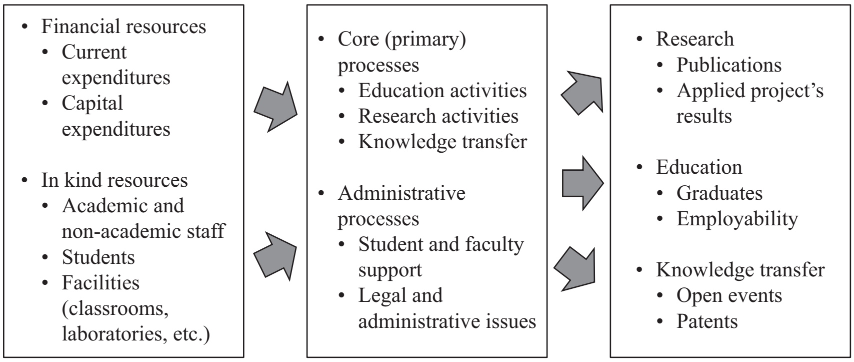
Tommaso Agasisti

1. INTRODUCTION: WHY IS STUDYING THE EFFICIENCY OF HIGHER EDUCATION INSTITUTIONS AND SYSTEMS AN IMPORTANT TASK?

Higher education (HE), under certain conditions, can play a key role in fostering the economic development of countries, regions and cities, especially when it helps in spreading innovation and improving human capital (De Meulemeester & Rochat, 1995; Lendel, 2010; Hanushek, 2016; Barra & Zotti, 2017; Leyden & Link, 2017; Valero & Van Reenen, 2019). The positive effects of HE, which spill over beyond the single students attending the educational programmes or the public and private entities engaged in the research activities, provide the key economic justification for the public funding of this sector (Toutkoushian & Shafiq, 2010).[1] The positive externalities of HE, in this perspective, are of different kinds: (i) a more educated workforce is more productive and more involved in the cultural and civil life, (ii) higher-level research contributes to new discoveries and inventions and (iii) outreach initiatives help the diffusion of innovation and the socialization of new knowledge (Wolfe & Haveman, 2001; McMahon, 2004; Chapman & Lounkaew, 2015).

Observing this field with economic lenses, the specific results of HE are pursued within organizations for which the primary mission is the production of teaching, research and knowledge transfer – that is, the Higher Education Institutions (HEIs). From an organization theory standpoint, the production process of these organizations can be modelled in a simple form: they employ financial and human resources (current expenditures, academic and non-academic staff, laboratories, students and so on) – which are usually strongly subsidized with public money – in well-structured processes (educational, research and administrative) to produce the desired levels of output – such as graduates, publications, patents and so on – see Figure 12.1 for a graphical representation.

According to the standard microeconomic theory, the single HEI acts in an efficient way, by minimizing its costs while maximizing the output produced. In other words, each HEI should be able to produce as many graduates, publications and knowledge transfer initiatives as possible, while keeping the costs associated with staff and structures at a minimum. There are, therefore, many reasons that suggest this is not (always) the case; as a consequence, the measurement of productivity of educational institutions is an important empirical and conceptual task (Hanushek & Ettema, 2017) and is still attracting an important research effort (Arias-Ciro, 2020). We can classify the inherent obstacles to efficiency, specifically in the HE sector, into three groups. First, many HEIs are public institutions and are subjected to the traditional sources of public sector inefficiency in production: soft-budget constraint, lack of incentives and so on (Wamsley & Zald, 1973). Second, HE is a typical field affected by the cost-disease dynamic, where the price of key inputs (especially academic staff and research laboratories)



Note: The indicators included as measures for inputs, processes and outputs are non-exhaustive examples, and the aim is only to clarify the broad model to the readers (these indicators are among those actually used in the literature about the efficiency of higher education institutions – see De Witte & Lopez-Torres, 2017).

Source: Author's elaboration.

Figure 12.1 The production process of HE within HE institutions – a reference model

increases at a higher rate than productivity (Archibald & Feldman, 2008; Martin, 2011). Third, the internal organization of universities and other HEIs is not always hierarchical, in other words it is not possible to coordinate the efforts of individuals and subunits towards the best practices and/or common goals (including efficiency efforts). In addition, there is evidence that management matters, and, for example, better managed departments are more productive in the delivery of teaching and research (McCormack *et al.*, 2014). Often, universities are managed mainly by academic and not by management professionals, so there is not *a priori* assurance of effective management at play.

The possibility of inefficient HE production is a key problem for our economies and society. Employing more resources than necessary for the development of educational and research activities translates into waste of public resources, which could be more appropriately dedicated to alternative uses for welfare and public services. The academic literature in the public finance field strongly debates how public spending can be useful for contributing to economic growth (Chen, 2006; Afonso & Furceri, 2010), also underlining the potential trade-off between alternative destinations of available public resources. Moreover, for a given level of public resources allocated to HE, its efficient use could result in higher levels of education and research produced with positive effects on current and potential economic growth. The negative effects of inefficient spending, conversely, could harm the medium- and long-run economic development of countries and regions (via a reduction of innovation gains). Lastly, a suboptimal efficiency of spending could negatively affect the other public (positive) externalities associated with HE, such as the improvement of civic, democratic and cultural skills of the population (Glaeser *et al.*, 2007).

The opportunity of adequately funding the HE sector has been sustained in various political and institutional circles, and the evidence of its potential beneficial impacts corroborates this necessity. Therefore, policy makers must guarantee that resources that are allocated to the sector are spent efficiently, in other words, the outputs generated by HEIs are obtained with the lowest possible investment – or, conversely, the HEIs are able to maximize the outputs produced with the available resources. In this vein, the two topics – level of public funding and its efficiency of use – are strictly intertwined.

The remainder of this chapter is organized as follows. In §1, I present some data about public resources allocated to HE in recent years, in a comparative perspective; also, the imperative of 'doing more with less' is critically discussed. In §2, I describe the two main methodological approaches used in the literature for assessing the efficiency of HE systems and institutions – namely, the parametric and non-parametric approaches. Section 3 synthetizes some evidence coming from a selection of studies about the efficiency of HE systems and institutions, as conducted in various areas of the world – with the aim of presenting a synthetic view of the state-of-the-art of the empirical evidence in the field. Lastly, §4 concludes with suggestions for future research.

1.1 Data about Higher Education Spending in OECD Countries

In this subsection, I report two key data about governmental spending on higher education (see Table 12.1), providing also an idea of their evolution over time. In Panel A, there is the total spending per student in Purchasing Power Parity US dollars, while in Panel B there is the spending as a percentage of GDP. In both cases, data refer to the expenditure of the general government for all the (public and private) tertiary education institutions. I included all OECD countries for which at least two observations are available in the selected years (that is 1995, 2000, 2005, 2010 and 2016). The database from which information is extracted is the annual publication by OECD, named Education at a Glance.[2]

From the reading of Table 12.1, two main pieces of evidence can be commented on. First, there is a substantial heterogeneity across countries in the amount of resources invested in tertiary education – and, consequently, relevant differences in the amount of resources available for the institutions. The (public) spending per student (in 2016) varies between around $2500 (Greece) and more than $20 000 (Sweden and Norway), with an average value of $9300. These differences correspond to a minimum allocation of 0.4 per cent of GDP (Japan) and a maximum of 1.7 per cent (Norway), with an average value of 0.9 per cent. Such differences in the spending decisions likely reflect heterogenous policy priorities, features of higher education systems and overall countries' economic and public finance situation. Nevertheless, any evaluation of the efficiency of HE spending should keep adequately in mind these differences in overall resources invested in the sector. Second, in all of the countries the overall public spending in the HE field is substantial (various billions of dollars), so ensuring its efficient use is a crucial issue, as well as is the development of interventions and policies that can improve it. This policy attention is of utmost importance if we observe the dynamic of public spending in the sector. In most countries included in Table 12.1, the pattern observed is one of stagnant growth expenditure, and in some cases the proportion of GDP invested is even declining. An important distinction is useful here; indeed, in many developed economies the resources devoted to HE are actually stable or declining (see, for example, Canada, Australia, the United States and France), while many emerging economies are increasing investment in the HE sector (see the cases of Brazil, Chile and Mexico). These different patterns suggest that emerging countries must create a more empirical base for studying the efficiency of their (growing) expenditure in the field. At the same time, the developed countries must face the challenge of keeping their level of HE production at a satisfactory level, without counting on growing contributions from public budgets – a topic which is directly connected with the quest for efficiency, which I discuss in §1.2.

Table 12.1 Panel A

Total (direct) expenditure on educational institutions (all private and public) per full-time equivalent student in tertiary education (level 5−8 ISCED11) from general government

Country	1995	2000	2005	2010	2016
Australia		6238 USD	6436 USD	7584 USD	6391 USD
Belgium			10 902 USD	13 551 USD	14 963 USD
Chile	1131 USD	1159 USD	923 USD	1677 USD	2698 USD
Czech Republic	5605 USD	4469 USD	5304 USD	6025 USD	7257 USD
Germany			11 160 USD	14 636 USD	14 468 USD
Denmark	9775 USD	11 246 USD	14 378 USD	18 040 USD	
Spain	3298 USD	5101 USD	7731 USD	10 318 USD	8273 USD
Estonia	2559 USD	2349 USD	2607 USD	4476 USD	10 411 USD
Finland	9595 USD	10361 USD	11855 USD	16 476 USD	16 221 USD
France			9287 USD	12 241 USD	12 474 USD
Greece	3008 USD	3565 USD	6111 USD		2529 USD
Hungary	5104 USD	4752 USD	4333 USD	5969 USD	7089 USD
Ireland	4697 USD	8655 USD	8652 USD		9491 USD
Iceland		8475 USD	8750 USD	8148 USD	12 888 USD
Israel	6759 USD	6786 USD	5281 USD	5605 USD	6241 USD
Italy	4175 USD	5390 USD	5250 USD	6644 USD	7066 USD
Japan			4690 USD	5974 USD	5864 USD
Lithuania		2415 USD	2842 USD	4779 USD	5107 USD
Latvia		1610 USD	2297 USD	3043 USD	4864 USD
Mexico	3305 USD	3661 USD	4294 USD	5681 USD	5077 USD
Netherlands	8898 USD	10 042 USD	11 058 USD	12 329 USD	13 168 USD
Norway	10 117 USD	12 471 USD	15 268 USD	18 100 USD	20 467 USD
Poland		2091 USD	3532 USD	4937 USD	7064 USD
Portugal	3934 USD	4543 USD	6361 USD	6688 USD	6761 USD
Russia				4,403 USD	5,408 USD
Slovakia	4585 USD	4301 USD	4439 USD	4948 USD	7932 USD
Sweden	13 200 USD	14 032 USD	13 247 USD	17 962 USD	20 486 USD
United States	6066 USD	8521 USD	9920 USD	10 269 USD	10 428 USD

Notes: Authors' elaborations on OECD data (Education at a Glance, various years). When cells are empty it means that data are missing. All countries in the last edition of OECD Education at a Glance were selected, but I include only those for which we have three years of observations. All data are reported in USD PPP (Purchasing Power Parity).

Table 12.1 Panel B

Total (direct) expenditure on educational institutions (all private and public) in tertiary education (level 5−8 ISCED11) from general government, as a percentage of GDP

Country	1995	2000	2005	2010	2016
Australia		0.75	0.69	0.78	0.75
Belgium			1.08	1.20	1.23
Brazil	0.63	0.67	0.75	0.86	1.00
Canada		1.49	1.34	1.58	1.25
Switzerland	1.05	1.05	1.27	1.16	1.27
Chile	0.37	0.38	0.30	0.53	0.80
Czech Republic	0.59	0.63	0.77	0.89	0.69
Germany			0.88	1.01	1.01
Denmark	1.59	1.48	1.60	1.69	
Spain	0.73	0.83	0.85	1.02	0.81
Estonia	0.97	1.02	0.78	1.04	1.24
Finland	1.84	1.60	1.59	1.77	1.55
France		1.10	1.07	1.19	1.12
Greece	0.55	0.74	1.41		0.62
Hungary	0.78	0.85	0.85	0.82	0.67
Ireland	0.86	1.10	0.90	1.17	0.57
Iceland		0.94	1.08	1.03	1.12
Israel	1.04	0.99	0.84	0.84	0.79
Italy	0.59	0.63	0.61	0.63	0.54
Japan			0.46	0.51	0.42
Lithuania		0.91	0.85	1.08	0.73
Latvia	0.68	0.69	0.76	0.71	0.68
Mexico	0.65	0.68	0.79	0.94	0.94
Netherlands	1.16	0.97	1.03	1.12	1.14
Norway	1.81	1.61	1.68	1.56	1.78
Poland	0.80	0.70	1.16	1.01	0.93
Portugal	0.82	0.87	0.84	0.91	0.71
Slovak Republic	0.69	0.68	0.68	0.62	0.69
Slovenia			0.94	1.02	0.84
Sweden	1.36	1.35	1.31	1.45	1.36
Turkey			0.74	0.90	1.41
United States		0.94	1.02	1.07	0.85

Notes: Authors' elaborations on OECD data (Education at a Glance, various years). When cells are empty it means that data are missing. All countries in the last edition of OECD Education at a Glance were selected, but we include only those for which we have three years of observations. All data are reported in percentage points (over the GDP).

1.2 'Doing More with Less' – An Imperative for Higher Education in the 2000s

Explicit attempts to improve efficiency have not been a recurrent topic in HE debates. Starting from the late 1990s and early 2000s, the decline in public funding for the sector in many countries imposed a discussion about the best way to efficiently use resources. 'Doing more with less' became an imperative in the field, following the direction taken by the public sector reforms under the New Public Management framework (Ferlie *et al.*, 1996).

In the specific HE context, such a policy approach concentrated on certain lines. The dynamics of change that happened in the HE sector are related to major trends and interventions in the funding system of HEIs, and specifically with the design and mechanisms of public funding (Liefner, 2003; Docampo, 2007; see also Chevaillier & Eicher, 2002 for a critical discussion of the European case). The main actions taken to impose a 'doing more with less' mentality in the HE sector can be grouped in three main categories:

- *Reduction of current expenditures*, by stimulating the efficient use of available resources. In practical terms, the definition of tight public budgets requires HEIs to seek a reduction in average costs for both education and teaching. All around the world, the proportion of public resources devoted to HE stagnated – or even reduced – so that the institutions cannot count on growing available resources granted by public authorities. With the same amount of money, the institutions must find a way to maintain the same level of activity, also because compressing some cost items (such as personnel) is often very difficult, especially for institutions belonging to the public sector. This approach calls the effectiveness of management into question, a typical suggestion coming from New Public Management (NPM) indications – namely, that public managers must hold strong managerial skills and competences for improving the efficiency of the organizations' operations (Romzek, 2000).
- *Searching for alternative revenues*, stimulating market-like mechanisms of governance to attract private resources (for example, students' fees and grants from companies), under the assumption that private stakeholders are more able to impose on HEIs a more efficient use of resources. This way, the overall costs of HE provision would be split across several partners and actors – a phenomenon known as cost sharing (Johnstone, 2004) – as well as the risks and responsibilities for monitoring the employment of resources.
- *Creating funding mechanisms that reward performance*, for example allocating public resources according to a set of performance metrics. Performance-based funding (PBF) soon became a popular approach, inspiring funding reforms in HE systems across the globe, especially for the component rewarding research productivity (Hicks, 2012). Despite the academic literature about public management raising concerns about the use of PBF when not properly considering contexts and information systems (Rabovsky, 2014 discusses this topic in the HE setting), there are theoretical arguments that support its use within the NPM theories. The most intriguing feature for supporters of 'doing more with less' philosophies is that an explicit reward for performance levels will induce the organizations to make the most of the available resources, with the aim of maximizing 'value for money' by attracting as much public funding as possible.

Together with understanding the best way to maximize the efficiency of spending in the sector, by means of using adequate funding mechanisms, the academic and institutional debate also considered how different funding schemes can have an effect on equality (that is, favouring the access of disadvantaged students) and performance (that is, stimulating the production of high-quality teaching and research) – see the discussion proposed by Barr (2004). Nevertheless, the discussion around the specific mechanisms for allocating funds does not reduce the attention of the various actors to the necessity of using the available resources in the most efficient fashion, making the topic of HE efficiency constantly at the centre of the debate and policy interest.

2. ASSESSING THE EFFICIENCY OF HIGHER EDUCATION: METHODOLOGICAL NOTES

In the academic literature about the empirical measurement of efficiency, the objective is to define a ‘frontier’ of efficient units (institutions) which is able to produce the highest level of output possible with the available resources (or, conversely, to minimize the use of resources for producing a given level of output). Empirically, data about inputs and outputs are used for deriving efficiency scores. The universities that obtain the best scores are classified as ‘efficient’ and a frontier of efficiency is drawn identifying the various combinations of inputs and outputs that can be regarded as efficient. From a methodological viewpoint, there are two broad categories of techniques that can be used for deriving this frontier of efficient HEIs (see Johnes, 2004; De Witte & Lopez-Torres, 2017; and Johnes, 2020 for a detailed discussion).

The first category is that of *parametric approaches*, rooted in the economic and econometric traditions, that specify the efficient production function of the institutions based on microeconomic theory. In this case, the coefficient of relationships between inputs and outputs is derived parametrically, and the deviation from the frontier is assumed to be caused partly by statistical (random) noise and partly by inefficiency in production. The first and most basic well-known method in this group is Stochastic Frontier Analysis (SFA), developed by Aigner *et al.* (1977). Formulating appropriate econometric hypotheses about the non-random component of the deviation from the frontier, an efficiency score can be computed for each institution, representing the eventual distance from the parametrically defined production function. From a practical perspective, the mathematical form of the production (or cost) function must be specified, according to the underlying economic theory. Especially when using the cost function as formulation (for coherence with the microeconomic theories of production), the problem is related to modelling the production of several outputs at the same time. Moreover, outputs are produced with joint resources – for example, the time and energy academic staff contribute to production of teaching and research simultaneously. These technical problems about the estimation of costs for multiproduct organizations have been faced since the early academic literature in the field of higher education economics (Cohn *et al.,* 1989). In the context of using multiproduct cost functions, the most frequently used forms employed in the HE field are Coob-Douglas, quadratic and translog (see Cohn & Cooper, 2004). In all cases, the specific cost functions selected for representing the (multiproduct) economic activity of universities must respect the methodological requirements indicated by Baumol *et al.* (1982).

The alternative category of approaches for measuring the efficiency of HEIs is that of non-parametric methods. In this case, the relationship between inputs and outputs is not defined parametrically, but instead by 'enveloping' input and output data observed from real institutions. Having defined the efficiency as the ratio of (weighted) outputs over (weighted) inputs, those units which are able to maximize such a ratio are deemed to be efficient. The frontier of efficient units is used as the benchmark for measuring the performance of the other units – the larger the distance between a unit and the frontier, the higher is its inefficiency. The weights assigned by linear programming techniques to each unit are those that maximize efficiency and, at least in theory, they reflect the different strategy adopted by each institution (that is, the inputs or outputs to which the decision makers assign stronger importance). The most frequently applied methodology in this area is Data Envelopment Analysis (DEA), which has been proposed by Charnes *et al.* (1978).

The two families of methods have strengths and weaknesses, which compensate each other. Parametric techniques (like SFA) hold the key advantage of deriving the efficiency scores in a statistically robust way, so that statistical inference can be applied to explore factors potentially associated with (in)efficiency. Indeed, starting from the model proposed by Battese and Coelli (1995), a variety of econometric techniques have been employed to model the 'determinants' of efficiency scores. These models have been extensively applied in the HE sector – see, for example, Robst (2001) for an early application to the case of the United State and Stevens (2005) for the case of England and Wales. These methods are particularly useful when exploring whether certain features of HEIs are associated with their efficiency or not – for example, the composition of faculty body, the composition of available funds, the degree of autonomy and so on. The main limitation of SFA is that it relies on two key assumptions: (i) the functional form of the production (cost) function is known, and (ii) the distribution of efficiency scores is also known. Both assumptions are very difficult to demonstrate and can be questioned on theoretical grounds. In particular, while the functional form for the production (cost) function is derived from standard microeconomic theory of organizations, there are good reasons to believe that HEIs often deviate from the hypothesized behaviour of standard economic theory – for example, there is no assurance of a cost-minimization incentive. If the actual (strategic) behaviour of universities cannot be modelled perfectly, it seems more reasonable to infer it by means of a non-parametric approach, that is, letting the data speak for themselves. This is the key advantage of using DEA (and its derivations) for analysing the efficiency of universities, and more generally of all those public or not-for-profit organizations which are not pursuing profit maximization (Emrouznejad & Yang, (2018). The flexibility of DEA can model production and cost functions of universities without the necessity of imposing strict functional forms, and so avoids the acceptance of very difficult theoretical assumptions. In addition, DEA can easily accommodate the specification of various inputs and outputs simultaneously, and this feature allows the method to model the production/cost function much more precisely considering the potential interactions across them. These advantages come at a cost, naturally. In its simplest formulation, DEA considers all of the deviation from the efficiency frontier as a source of inefficiency in a very determinist way. In other words, the method does not consider the potential for measurement errors and/or for statistical noise. This is a strong limitation, especially because data about university budgets and activities are often measured imprecisely. The development of bootstrap-DEA (Simar & Wilson, 2000) only partially solves the problem of the statistical validity of DEA efficiency estimates, therefore the deterministic nature of the method remains a strong limitation.

Given the relative strengths and weaknesses, preferring one or another approach is not straightforward and definitive; for this reason, the existing contributions employ both methods in different circumstances and examples. Only a few studies attempt to compare the results obtained by using the two methods alternatively (on the same set of data) in the HE field, as for example McMillan and Chan (2006), Agasisti and Dal Bianco (2009) and Johnes (2014). The scarce use of such comparisons is reasonable given the widely different assumptions behind the different methodological approaches.

It is important to remark that SFA and DEA are the two basic approaches that have been improved and developed in decades of methodological and empirical research; the interested reader who wants to know more about recent advancements in these methodologies – as used in the educational context – can refer to the excellent essay by Johnes (2020). Anyway, the current academic literature still refers to the basic models for most empirical analyses of tertiary education institutions (see the review by De Witte & Lopez-Torres, 2017).

3. WHAT WE KNOW ABOUT THE EFFICIENCY OF HIGHER EDUCATION: A CRITICAL READING OF RECENT, EXISTING LITERATURE

Realizing a comprehensive analysis of what emerges from the wide academic literature about the efficiency of universities and HE systems is a complex task, which is well beyond the scope of this chapter. There are dozens of papers published in the past 20 years, which undertake assessments of HEIs' efficiency in a variety of countries; notably, most of these studies refer to European countries (mainly the United Kingdom, Spain, Germany, Italy), Australia, the United States and China. The interested reader can have a first, synthetic view of this field through the reviews in Worthington (2001), Salerno (2003), Johnes, (2004) and De Witte and Lopez-Torres (2017). The overall picture that emerges from this (very heterogenous) literature is twofold: (i) the average efficiency of universities is relatively high (the average degree of inefficiency reported in most HE systems is between 10 per cent and 15 per cent), and (ii) it is possible to identify specific factors associated with efficiency, such as for the example the composition of the faculty, the subject mix between humanities, social science and scientific fields, the sources of revenue, size and so on. This way, policy makers at the national level can make detailed suggestions about levers that can be activated or aspects that can be modified for improving the average efficiency of the universities operating in their countries. At the same time, managers of single institutions can have a benchmark to compare their operations and activities. Therefore, although country-specific studies are very informative about situations in single systems, it is very difficult to draw definitive views about the overall efficiency of the universities around the globe.

Given that previous studies already provide a comprehensive picture of the literature that concentrates on country-specific assessments of universities' efficiencies, the main aim of this section is to describe the findings from some recent contributions, tackling a specific angle which is the comparison of efficiency scores for different HE systems or universities, that is, an international perspective. Such a specific focus has two main advantages. First, by considering only more recent contributions the chapter discusses recent data, which can be interesting for analysts and policy makers. Second, by considering the international perspective, it contributes to a wider policy horizon, implying messages and stimuli for an international readership. The section includes studies which consider two different levels of analysis: whole HE systems (national level) or single HEIs in different countries.

Few papers consider the entire HE systems as units of analysis.[3] St. Aubyn *et al.* (2009) assess the efficiency of 23 European HE systems, in addition to those of the United States and Japan, for the period 1998–2005. Semi-parametric methods and SFA are both employed, and the findings identify a group of efficient HE systems across a variety of model specifications (Ireland, Japan, Sweden, the United Kingdom and the Netherlands). Some other countries report very low efficiency scores (such as Bulgaria, Spain, Hungary, the Czech Republic, Slovakia, Estonia, Portugal and Greece). The authors are interested in exploring the system-level features associated with higher levels of efficiency, and indicate the following elements: a well-performing secondary system, output-based funding rules (that is, incentives towards performance), institutions' formal and independent evaluation and institutions' autonomy in staff policy. Agasisti (2011) performs a similar exercise, with the notable difference that the focus is on 18 European HE systems, it considers only the period 2000–03 and DEA is adopted as the methodology. The HE systems which are identified as efficient are those of Switzerland and United Kingdom – but with less differences than those reported by St. Aubyn *et al.* (2009). Among the factors associated with efficiency, GDP per capita shows a positive correlation with efficiency, while public funding a negative one; this latter result is particularly important from a public finance perspective, as it suggests the potential room for containing public spending without harming the production of HE outputs. Unfortunately, there are no recent studies that provide updated measures of efficiency at the country level in a dynamic (panel) perspective; this is a gap that must be addressed in the next few years, as more recent data can support positive policies and actions at the supranational level. The paper by Kosor *et al.* (2019) assesses the efficiency of HE systems of the EU-28 countries for the single year, 2015, using a DEA method. The results indicate that the average efficiency level is around 90 per cent, with many HE systems appearing to use their public funding efficiently. Nevertheless, some countries obtain very low efficiency scores, in particular Greece and Italy.

The earliest attempt at comparing the efficiency of single universities across more than one country comes from the innovative work by Joumady and Ris (2005). The authors exploit the available data obtained through interviews conducted with graduates from more than 200 universities in eight European countries. The efficiency scores derived are used for ranking single institutions, and for obtaining system-level average performance. However, the bulk of the (still scarce) academic literature which studies the efficiency of universities in a cross-national perspective utilizes a relatively new dataset developed by the European Commission, named ETER[4] – European Tertiary Education Register, and also the previous versions of it called Aquameth and Eumida.[5,6] By means of these new data, analysts can explore the production processes of universities across different countries within the European boundaries, being substantially sure that variables and their measurement are sufficiently comparable across the HE systems. Daraio *et al.* (2015) use an extract from the database for the year 2008, including 400 universities in 16 European countries, and apply a robust non-parametric efficiency technique. The results suggest that size and subject specialization are statistically associated with the efficiency of universities in the various countries, although efficiency in research activity benefits from multidisciplinary interactions. Bolli *et al.* (2016) employ data of universities from eight European countries for the period 1994–2006, with the aim of evaluating whether different types of funding affect their efficiency. Through the use of SFA methodology, they demonstrate that the use of different types of funding mechanisms (in particular, fees vs competitive research grants) can generate different effects on the efficiency frontier and scores of single universities. Veiderpass and McKelvey (2016) explore the efficiency of around

950 universities in 17 European countries, adopting DEA as the methodology. The findings indicate that the most efficient universities are located in countries which spend less on HE; moreover, the size of institutions and research intensity are positively associated with the efficiency scores in all the HE systems considered. Herberholz and Wigger (2020) perform a DEA efficiency analysis of 450 universities across 16 European countries, for the period 2011–14. Their approach starts with defining clusters of homogenous universities, according to their discipline profile (for differentiating between humanities, social sciences and scientific subjects). The results indicate that universities in some countries are more efficient, on average (Poland, Belgium and the Netherlands). When considering the factors associated with efficiency, the study suggests positive correlations with the share of full professors, the proportion of external (non-public) funding and size (as measured through the number of students). Importantly, the authors claim that fair comparisons of efficiency must take into account structural differences due to different subject mixes, otherwise these differences would be improperly attributed to (in)efficient use of resources. This recommendation is in line with the study by Bonaccorsi and Daraio (2008), who clearly indicate that universities actually pursue different strategies, so that their efficiency must be evaluated accordingly for taking such heterogeneity into account.

Wolszczak-Derlacz (2017) compares, for the first time, the efficiency of 500 universities located in 10 European countries and in the United States (using a double-bootstrap DEA methodology as prescribed by Simar & Wilson, 2007); this is a unique study that adopts a truly international perspective, outside the boundary of two different continents. The time window under scrutiny is the decade between 2000 and 2010. The findings presented in this work are particularly relevant, as they constitute the state-of-the-art knowledge about the relative efficiency of the HE systems. When assuming a cross-continental frontier, the mean value of efficiency scores is very similar in Europe and in the United States, indicating that universities on both continents can substantially increase their outputs (graduates and publications) by around 35 per cent – on average – without employing more resources. Of course, these average results mask great heterogeneity within both European and US HE systems; moreover, US universities report much higher spending per capita, implying that they also produce more graduates and publications (although at a similar efficiency rate as their European counterparts, that is, producing a higher volume of outputs). When considering the factors associated with efficiency, the most notable evidence is that public funding tends to have a negative correlation with efficiency at European universities, while the same relationship does not hold for the US institutions. This element stimulates reflections for public finance policies, indicating that the dependence on public budgets can be detrimental under specific circumstances but not in others. Additionally, in both HE systems the most efficient universities are located in the most economically developed regions, suggesting that the efficiency of operations is somehow influenced by the economic activities of the territory where the universities work.

Much less evidence is available about the relative efficiency of HEIs in other areas of the world, such as Africa, Asia or Latin America – although some recent studies report some first empirical explorations of the universities' efficiency in those areas (see Munoz, 2016; Myeki & Temoso, 2019; Villano & Tran, 2019). For other areas, like China and Australia, there are many more country-specific studies of universities' efficiency (see, for example the recent contributions by Yang *et al.*, 2018 and Lee & Worthington, 2016), but universities from these countries are not included in more international benchmarking exercises. As pointed out in §4, the opportunity of building a more comprehensive view of the international frontier of efficient universities and HE systems is one fundamental challenge for future research in the field.

4. THE WAY FORWARD: LINES FOR FUTURE RESEARCH IN THE FIELD

In the light of the main aspects emerging from the current discussion about the efficiency of public spending in higher education, it is worth indicating three main lines for future research in the field. They should be conceived as additional to the ongoing effort of studying efficiency of HE systems and institutions. My opinion is that a constant, perseverant research agenda on efficiency levels of HE in various countries is necessary for both academic and institutional reasons. Indicating lines for research development in the area is coherent with the nature of this Handbook, which questions current and future trends in the broader Public Finance field.

4.1 Cross-country Comparisons of HE Efficiency – Identifying the International Efficiency Frontier

In §3, a description of current recent studies confirmed the ongoing interest in exploring the efficiency of HE systems, as well as the efficiency of single institutions across countries. Such an interest is understandable, given the potential for inspiring policies and actions from different contexts, by learning from the experience of various countries – see, for example, the well-known discussion about what European universities can learn from the US HE system, as proposed by Aghion *et al*. (2010).

This strand of academic research necessitates some further development, however. A present limitation is that some of the recent attempts at comparing HE systems in very different areas of the world systematically consider only two of them at the same time, as in Wolszczak-Derlacz (2017). In these empirical exercises, other areas of the globe should be included, such as Australia, China and so on, where very well-developed HE systems exist. A full picture of the international efficiency frontier of universities should consider all these areas simultaneously, a challenging task which requires some investment in building comparable and credible datasets.

When considering the potential benefits coming from international benchmarking, however, it must be acknowledged that efficiency studies in some countries and areas of the world are less frequent than in others. For instance, Arias-Ciro (2020) underlines how few contributions exist about Latin American countries. Moore *et al.* (2019) take the initiative to systematically explore the efficiency of universities and other HEIs in a group of Asian countries, like Cambodia, Pakistan, Thailand and so on – which is an area of the world where the number of studies about HE efficiency is very limited. Given the steady growth in public resources devoted to HE in these countries, a clear direction for future research consists of producing more evidence about the efficiency of HEIs in this area, also comparing the obtained results with those in different countries – this way deepening the understanding of the international frontier of efficient public spending in the field.

4.2 Determinants of HE Efficiency – Moving Towards Exploring Causal Effects

A number of studies demonstrate that the (in)efficiency of Higher Education Institutions is statistically correlated to a number of their characteristics, such as the size of the institution (related to the exploitation of economies of scale), the number and composition of faculties, the sources of funding and gender staff composition (for example, Wolszczak-Derlacz & Parteka, 2011). The analyses aiming at detecting associations with efficiency scores usually rely upon two-stage procedures (in the case of non-parametric methods) or upon statistical modelling of inefficiency components (in the case of parametric approaches). In both cases the coefficients of correlation between efficiency scores and the selected characteristics must be interpreted as statistical associations, and not as causal influences. For exploring causal effects on efficiency, a different framework is needed. As pointed out by De Witte and Lopez-Torres (2017), this is an important challenge for scholars who want to move a step forward to the use of efficiency studies for supporting and suggesting policies in the HE field.

There are various promising avenues for introducing a more formal and robust assessment of causality within efficiency studies in the Higher Education field. For example, Santín and Sicilia (2018) exploit a natural experiment for assessing the effectiveness of a policy using an efficiency framework; the specific exercise refers to primary schools but can be similarly extended to HE. More generally, Santín and Sicilia (2017) suggest a general method for dealing with endogeneity in efficiency estimates that use Data Envelopment Analysis – where endogeneity is a major source of threat to causal inference. Also, the combination of adequate econometric techniques (such as Instrumental Variables or matching procedures) with efficiency analyses can better identify causal effects of variables on efficiency scores, by controlling for selection bias.

Without further entering into the technical details of various ways to improve the analyses of causal effects on efficiency, it is important to highlight this as an area of important development for the research in the field. Indeed, the provision of findings that identify clear and robust causal effects on efficiency can serve the decision makers – especially managers of single institutions and national policy makers – with actionable indications and suggestions for improving the value for public money invested in the HE sector.

4.3 Tackling the Quality Issue – Measuring the Quality of Output and Efficient Production

In comparing the production processes of universities, the analyst assumes that the quality of outputs is homogenous across the various institutions – this way, the efficiency can be defined as the ability of maximizing the (quantity of) outputs given the available (quantity of inputs). Such an assumption – often left implicit in the academic papers – is, however, hard to be accepted silently. There are two major threats to the validity of this key assumption:

- The quality of available inputs is different (for example, the skills and competences of students when they enrol at the university) or they are affected by external conditions that influence their use in the production process – a phenomenon which Johnson and Ruggiero (2014) call 'environmental harshness' with specific reference to education.
- The quality of the produced outputs is heterogenous across universities, so the pure measurement of quantities does not allow considering inefficiencies in the production processes (as they are masked by quantities, indeed).

Both these threats are serious and can impair unbiased comparisons of universities' efficiency. Taking into account these quality differences is challenging, albeit possible. A few examples can help here. Data about students (typically, used as inputs in efficiency analyses) can be 'corrected' for considering the measured level of competences at the enrolment (this is the approach used by Agasisti & Salerno, 2007). On the output side, the number of academic publications (a typical output in efficiency analyses) can be weighted through citations to measure their impact/quality (see Abbott & Doucouliagos, 2003). None of these methods are perfect; nevertheless, they are valuable attempts at dealing with a fundamental issue in efficiency measurement. Future research should be devoted to developing protocols and guidelines for integrating the consideration for input/output quality into standard efficiency measurement and comparisons across institutions. In this vein, Daraio *et al.* (2019) pave the way for concrete advancements, by suggesting a method for identifying latent quality variables within efficiency analyses.

NOTES

1. See also an interesting discussion in Winter and Pfitzner (2013).
2. I employed several years/editions of Education at a Glance for analysing the relevant data. I am grateful to Eng. Alice Bertoletti for the help in gathering the data and creating the tables presented in this section. Please refer to the OECD website for further details about the way in which statistics are computed: https://www.oecd.org/education/education-at-a-glance/
3. The paper by Williams *et al.* (2013) describes the performance of 48 HE systems across the globe, using 20 variables classified in specific macro-categories. However, the analysis does not present an efficiency evaluation of the systems themselves.
4. For details about the ETER Project and the current content of this dataset, please visit the website https://www.eter-project.com/#/home
5. It is worth recalling that another interesting stream of the literature is the one comparing the efficiency of universities in (only) two different countries. For example, Italian universities' efficiency scores have been compared with those of their English (Agasisti & Johnes, 2009), Spanish (Agasisti & Perez-Esparells, 2010), Dutch (Agasisti & Haelermans, 2016) and Polish (Agasisti & Wolszczak-Derlacz, 2015) counterparts.
6. An alternative, interesting approach is presented by Bolli (2011); in this (unpublished) paper, the author uses the QS World Ranking as data source for deriving efficiency scores of 273 top research universities across 29 countries in the period 2007−09.

REFERENCES

Abbott, M., & Doucouliagos, C. (2003). The efficiency of Australian universities: a data envelopment analysis. *Economics of Education Review*, 22(1), 89–97.

Afonso, A., & Furceri, D. (2010). Government size, composition, volatility and economic growth. *European Journal of Political Economy*, 26(4), 517–532.

Agasisti, T., & Salerno, C. (2007). Assessing the cost efficiency of Italian universities. *Education Economics*, 15(4), 455–471.

Agasisti, T., & Dal Bianco, A. (2009). Measuring efficiency of higher education institutions. *International Journal of Management and Decision Making*, 10(5–6), 443–465.

Agasisti, T., & Johnes, G. (2009). Beyond frontiers: comparing the efficiency of higher education decision-making units across more than one country. *Education Economics*, 17(1), 59–79.

Agasisti, T., & Pérez-Esparrells, C. (2010). Comparing efficiency in a cross-country perspective: the case of Italian and Spanish state universities. *Higher Education*, 59(1), 85–103.

Agasisti, T. (2011). Performances and spending efficiency in higher education: a European comparison through non-parametric approaches. *Education Economics*, 19(2), 199–224.

Agasisti, T., & Wolszczak-Derlacz, J. (2015). Exploring efficiency differentials between Italian and Polish universities, 2001–11. *Science and Public Policy*, 43(1), 128–142.

Agasisti, T., & Haelermans, C. (2016). Comparing efficiency of public universities among European countries: different incentives lead to different performances. *Higher Education Quarterly*, 70(1), 81–104.

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., & Sapir, A. (2010). The governance and performance of universities: evidence from Europe and the US. *Economic Policy*, 25(61), 7–59.

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1), 21–37.

Archibald, R. B., & Feldman, D. H. (2008). Explaining increases in higher education costs. *The Journal of Higher Education*, 79(3), 268–295.
Arias-Ciro, J. (2020). Bibliometric study of the efficiency of public expenditure on education. *Revista CEA*, 6(11), 127–144.
Barr, N. (2004). Higher education funding. *Oxford Review of Economic Policy*, 20(2), 264–283.
Baumol, W., Panzar, J., & Willig, R. (1982). *Contestable Markets and the Theory of Industry Structure*. Harcourt Brace Jovanovich.
Barra, C., & Zotti, R. (2017). Investigating the human capital development–growth Nexus: does the efficiency of universities matter?. *International Regional Science Review*, 40(6), 638–678
Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20(2), 325–332.
Bolli, T. (2011). The global production frontier of universities. *KOF Swiss Economic Institute, ETH Zurich Working Paper*, (272).
Bolli, T., Olivares, M., Bonaccorsi, A., Daraio, C., Aracil, A. G., & Lepori, B. (2016). The differential effects of competitive funding on the production frontier and the efficiency of universities. *Economics of Education Review*, 52, 91–104.
Bonaccorsi, A., & Daraio, C. (2008). The differentiation of the strategic profile of higher education institutions. New positioning indicators based on microdata. *Scientometrics*, 74(1), 15–37.
Chapman, B., & Lounkaew, K. (2015). Measuring the value of externalities from higher education. *Higher Education*, 70(5), 767–785.
Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
Chen, B. L. (2006). Economic growth with an optimal public spending composition. *Oxford Economic Papers*, 58(1), 123–136.
Chevaillier, T., & Eicher, J. C. (2002). Higher education funding: a decade of changes. *Higher Education in Europe*, 27(1–2), 89–99.
Cohn, E., Rhine, S. L., & Santos, M. C. (1989). Institutions of higher education as multi-product firms: economies of scale and scope. *The Review of Economics and Statistics*, 71(2), 284–290.
Cohn, E., & Cooper, S. T. (2004). 15 Multi-product cost functions for universities: economies of scale and scope. *International Handbook on the Economics of Education*, 579.
Daraio, C., Bonaccorsi, A, & Simar, L. (2015). Efficiency and economies of scale and specialization in European universities: a directional distance approach, *Journal of Informetrics*, 9, 430–448.
Daraio, C., Simar, L., & Wilson, P. W. (2019). Quality and its impact on efficiency (No. 2019/06). *LEM Working Paper Series*.
De Meulemeester, J. L., & Rochat, D. (1995). A causality analysis of the link between higher education and economic development. *Economics of Education Review*, 14(4), 351–361.
De Witte, K., & López-Torres, L. (2017). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society*, 68(4), 339–363.
Docampo, D. (2007). International comparisons in higher education funding. *Higher Education in Europe*, 32(4), 369–386.
Emrouznejad, A., & Yang, G. L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*, 61, 4–8.
Ferlie, E., Fitzgerald, L., & Pettigrew, A. (1996). *The new public management in action*. Oxford University Press, Oxford.
Glaeser, E. L., Ponzetto, G. A., & Shleifer, A. (2007). Why does democracy need education? *Journal of Economic Growth*, 12(2), 77–99.
Hanushek, E. A. (2016). Will more higher education improve economic growth? *Oxford Review of Economic Policy*, 32(4), 538–552.
Hanushek, E. A., & Ettema, E. (2017). Defining productivity in education: Issues and illustrations. *The American Economist*, 62(2), 165–183.
Herberholz, L., & Wigger, B. U. (2020). Efficiency of european universities: A comparison of peers. *CESifo Working Papers,* 8044.
Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251–261.
Johnes, J. (2004). 16 Efficiency measurement. *International handbook on the economics of education*, 613.
Johnes, J. (2014). Efficiency and mergers in English higher education 1996/97 to 2008/9: parametric and non-parametric estimation of the multi-input multi-output distance function. *The Manchester School*, 82(4), 465–487.
Johnes, G. (2020). Economic approaches to school efficiency. In *The Economics of Education* (pp. 479-489). Academic Press.
Johnson, A. L., & Ruggiero, J. (2014). Nonparametric measurement of productivity and efficiency in education. *Annals of Operations Research*, 221(1), 197–210.
Johnstone, D. B. (2004). The economics and politics of cost sharing in higher education: comparative perspectives. *Economics of Education Review*, 23(4), 403–410.

Joumady, O., & Ris, C. (2005). Performance in European higher education: A non-parametric production frontier approach. *Education Economics*, 13(2), 189–205.
Kosor, M. M., Perovic, L. M., & Golem, S. (2019). Efficiency of public spending on higher education: a data envelopment analysis for EU-28. *Problems of Education in the 21st Century*, 77(3), 396.
Lee, B. L., & Worthington, A. C. (2016). A network DEA quantity and quality-orientated production model: An application to Australian university research services. *Omega*, 60, 26–33.
Lendel, I. (2010). The impact of research universities on regional economies: the concept of university products. *Economic Development Quarterly*, 24(3), 210–230.
Leyden, D. P., & Link, A. N. (2017). Knowledge spillovers, collective entrepreneurship, and economic growth: The role of universities. In *Universities and the Entrepreneurial Ecosystem*. Edward Elgar Publishing.
Liefner, I. (2003). Funding, resource allocation, and performance in higher education systems. *Higher Education*, 46(4), 469–489.
Martin, R. E. (2011). *The college cost disease: Higher cost and lower quality*. Edward Elgar Publishing.
McCormack, J., Propper, C., & Smith, S. (2014). Herding cats? Management and university performance. *The Economic Journal*, 124(578), F534–F564.
McMahon, W. W. (2004). The social and external benefits of education. *International handbook on the economics of education*, 211.
McMillan, M. L., & Chan, W. H. (2006). University efficiency: a comparison and consolidation of results from stochastic and non-stochastic methods. *Education Economics*, 14(1), 1–30.
Myeki, L. W., & Temoso, O. (2019). Efficiency assessment of public universities in South Africa, 2009-2013: panel data evidence. *South African Journal of Higher Education*, 33(5), 264–280.
Moore, K., Coates, H., & Croucher, G. (2019, March). Measuring international higher education productivity: Lessons from nine countries in Asia. In *Higher Education Forum,* 16, 69–84. Research Institute for Higher Education, Hiroshima University.
Munoz, D. A. (2016). Assessing the research efficiency of higher education institutions in Chile. *International Journal of Educational Management*, 30(6), 809–825.
Rabovsky, T. (2014). Support for performance-based funding: the role of political ideology, performance, and dysfunctional information environments. *Public Administration Review*, 74(6), 761–774.
Robst, J. (2001). Cost efficiency in public higher education institutions. *The Journal of Higher Education*, 72(6), 730–750.
Romzek, B. S. (2000). Dynamics of public sector accountability in an era of reform. *International Review of Administrative Sciences*, 66(1), 21–44.
Salerno, C. (2003). *What we know about the efficiency of higher education institutions: The best evidence* (Vol. 99). Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.
Santín, D., & Sicilia, G. (2018). Using DEA for measuring teachers' performance and the impact on students' outcomes: evidence for Spain. *Journal of Productivity Analysis*, 49(1), 1–15.
Simar, L., & Wilson, P. W. (2000). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics*, 27(6), 779–802.
Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1), 31–64.
St. Aubyn, M. S., Garcia, F., & Pais, J. (2009). *Study on the efficiency and effectiveness of public spending on tertiary education* (No. 390). Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.
Stevens, P. A. (2005). A stochastic frontier analysis of English and Welsh universities. *Education Economics*, 13(4), 355–374.
Toutkoushian, R. K., & Shafiq, M. N. (2010). A conceptual analysis of state support for higher education: Appropriations versus need-based financial aid. *Research in Higher Education*, 51(1), 40.
Valero, A., & Van Reenen, J. (2019). The economic impact of universities: evidence from across the globe. *Economics of Education Review*, 68, 53–67.
Veiderpass, A., & McKelvey, M. (2016). Evaluating the performance of higher education institutions in Europe: a nonparametric efficiency analysis of 944 institutions. *Applied Economics*, 48(16), 1504–1514.
Villano, R. A., & Tran, C. D. T. T. (2019). Technical efficiency heterogeneity of tertiary institutions in Viet Nam: a metafrontier directional technology approach. *Technological and Economic Development of Economy*, 25(6), 1058–1080.
Wamsley, G. L., & Zald, M. N. (1973). The political economy of public organizations. *Public Administration Review*, 12(1), 62–73.
Williams, R., de Rassenfosse, G., Jensen, P., & Marginson, S. (2013). The determinants of quality national higher education systems. *Journal of Higher Education Policy and Management*, 35(6), 599–611.
Winter, S., & Alexander, P. (2013). Externalities and optimal subsidization of higher education. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2281207.
Wolfe, B. & Haveman, R. (2001). Accounting for the social and non-market benefits of education. In John Helliwell (ed.), *The Contribution of Human and Social Capital to Sustained Economic Growth and Well Being*. Paris: OECD.

Wolszczak-Derlacz, J., & Parteka, A. (2011). Efficiency of European public higher education institutions: a two-stage multicountry approach. *Scientometrics*, 89(3), 887–917.
Wolszczak-Derlacz, J. (2017). An evaluation and explanation of (in) efficiency in higher education institutions in Europe and the US with the application of two-stage semi-parametric DEA. *Research Policy*, 46(9), 1595–1605.
Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics*, 9(3), 245–268.
Yang, G. L., Fukuyama, H., & Song, Y. Y. (2018). Measuring the inefficiency of Chinese research universities based on a two-stage network DEA model. *Journal of Informetrics*, 12(1), 10–30.

13. Efficiency of public health spending

Pedro Pita Barros and Eduardo Costa

1. THE PUBLIC HEALTH SECTOR: WHY IS IT SPECIAL?

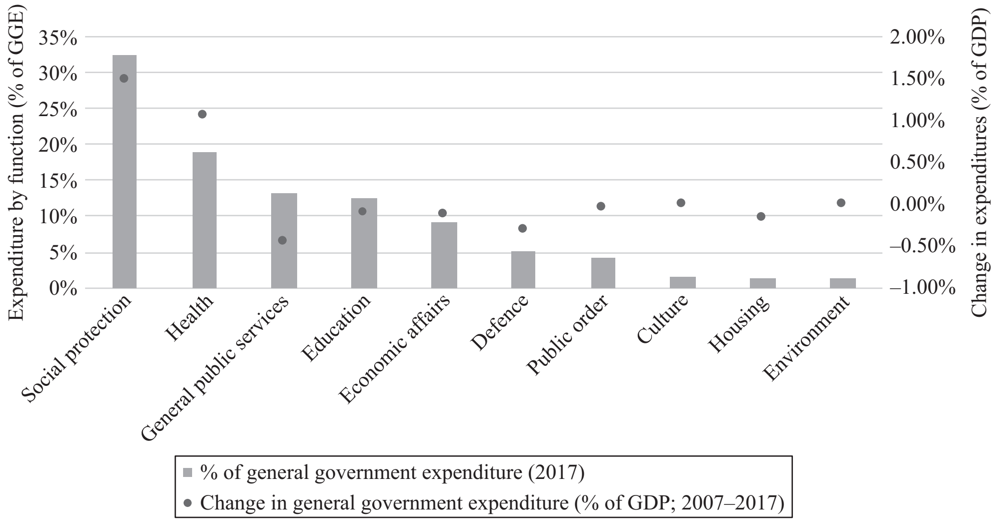
Access to high-quality care is a priority worldwide. The desire to protect citizens against adverse health shocks has motivated multiple and ambitious programmes across the world. The introduction of national social insurance or the provision of health care through a national health service are just some examples. In fact, the goal of achieving universal health coverage worldwide has become a major priority objective of the World Health Organization.

Despite the important role that private entities can play in funding or providing health care, attaining such an ambitious objective implicitly recognizes the scope for government intervention. Public health spending is related to the need to provide adequate health coverage and access to the population, as well as of reducing unmet needs. This implicit role attributed to the public health sector was amplified following the COVID-19 pandemic, even in countries where the public health sector plays usually a smaller role. The need for public health investment, for national and supra-national coordination, as well as the need to overcome market failures, create multiple grounds for government intervention in the health sector. Such intervention can take multiple forms, from regulation, financing or direct provision of health care. Regardless of the form taken, it is clear that public health spending is a crucial part of modern health systems.

The unequivocal importance of the health sector to modern societies, alongside the rationale for public intervention, led the public health sector to gain a main role in general government expenditure. In 2018, general government expenditure represented 41 per cent of GDP, on average, for high-income OECD countries (OECD, 2019). Health spending is one of the main areas of government intervention included, representing 19 per cent of public spending, surpassed only by social protection spending. The importance of the public health sector, measured by its proportion of public expenditure, has also increased over time. It is the sector with the second highest growth rate in government spending in the decade 2007–17 (Figure 13.1), representing an increasing share of GDP.

Total health spending can be funded by compulsory health insurance, voluntary health insurance and out-of-pocket payments. The following figure displays the composition of health spending for a set of European countries[1] (Eurostat, 2020).

Compulsory health insurance is mandatory expenditure that aims at protecting individuals against negative health shocks. On average, it represents 76 per cent of total health spending, and it can be either public (41 per cent) or private (36 per cent). Compulsory public insurance (which is the focus of this chapter and will be referred to as 'public health spending') is determined by the government, who has the direct responsibility and a specific budget to manage the health programme. On the other hand, compulsory private spending represents health care protection which is determined by law or by the government based on the payment of contributions by or on behalf of individuals (for example social health insurance, mandatory health insurance and so on).



Source: OECD; own calculations. Selected OECD countries.[2]

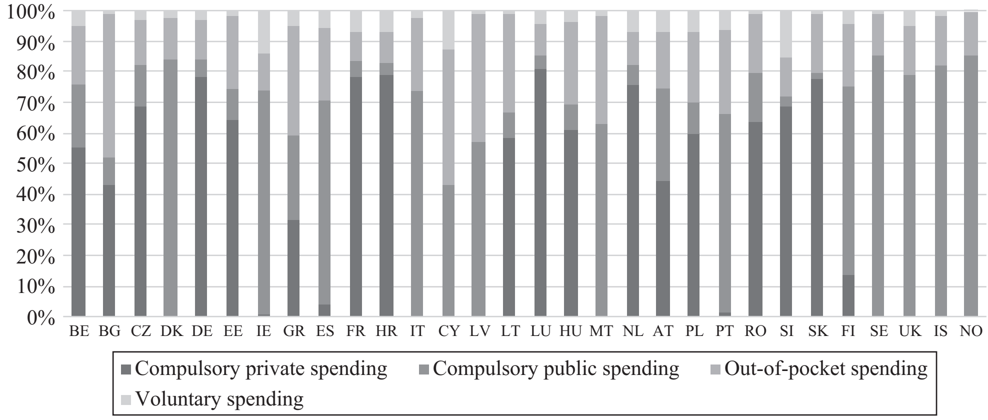
Figure 13.1 Expenditure by function (per cent of general government expenditure; left axis; 2017) and change in expenditures by function (per cent of GDP; right axis; 2007−17)

Voluntary health insurance (5 per cent of total health spending) represents spending by individuals that buy additional health insurance, to achieve higher levels of protection, even if not mandated by the government.

Out-of-pocket payments (19 per cent of total health spending) are direct health care payments made by individuals at the point of use. Still, such payments can result from the absence of insurance protection (for instance, if a patient goes to hospital without insurance and has to pay the full price), or from user charges determined by the insurance protection (for instance, co-payments mandated by the insurance firm). Thus, part of these payments is determined by insurance, while the remaining represents a direct decision of the individual to pay for immediate use of health care services or products. However, data in Figure 13.2 does not allow us to decompose those two effects.

Health spending results from a price and a quantity effect. Higher health spending growth does not necessarily imply the provision of more or better health services. Instead, such growth might reflect cost increases on some of those services. By the same token, cost containment measures might achieve price reductions – which will be reflected in lower health spending, even if the level of health services remains unchanged. Also, health spending levels do not account for debt contracted by the health sector not paid in the current year. By not considering debts, the required spending level for the observed health outcomes would be greater than that registered for that particular year. One should be careful when analysing health spending data given that the price and quantity effects might not be immediately or clearly displayed by the data.

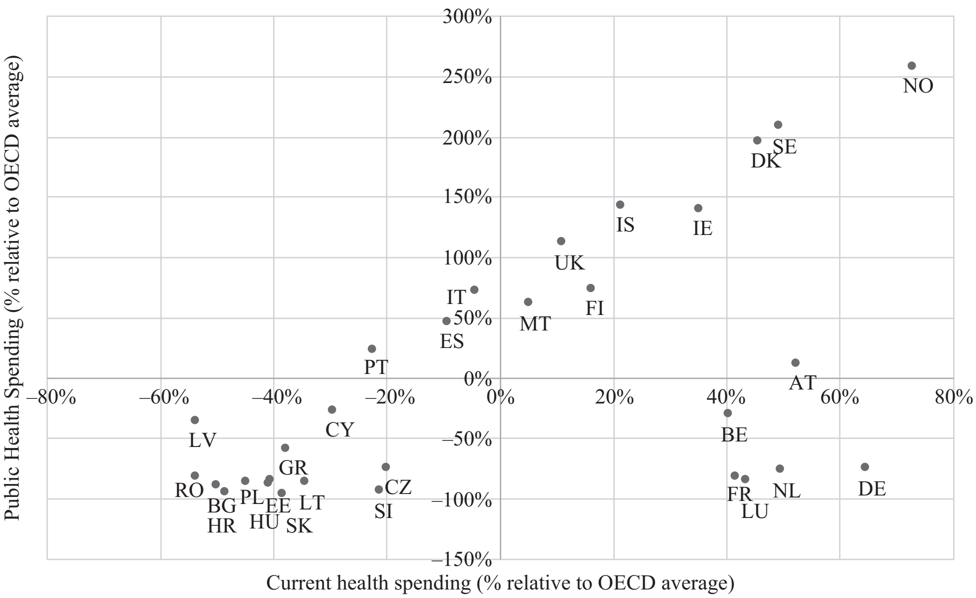
The importance of public health spending in total health spending varies considerably across countries. Figure 13.3 represents such variables relative to the European average (Eurostat, 2020). The European average is used as a benchmark to aggregate countries, but it does not



Source: Eurostat; 2018 (or closest year); own calculations.

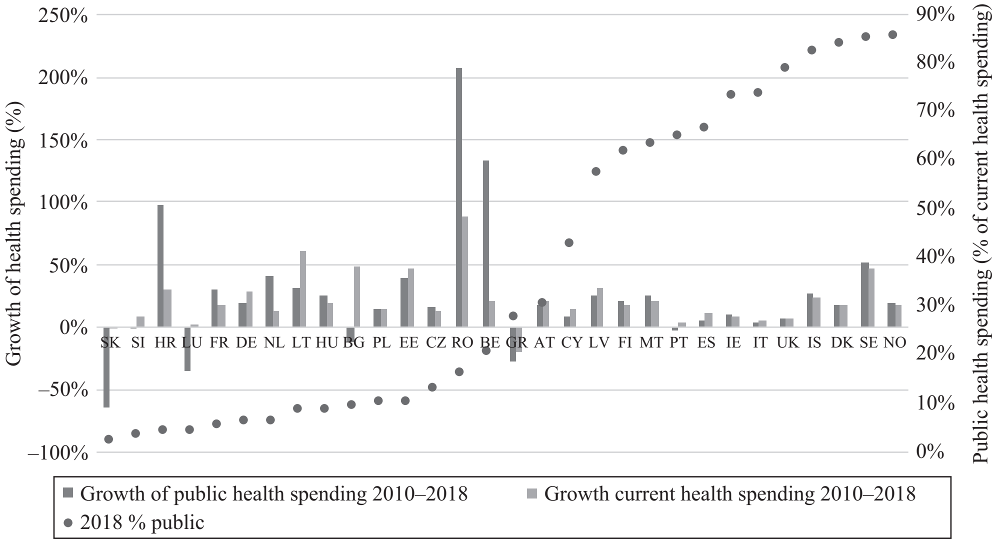
Figure 13.2 Decomposition of current health spending (per cent; adds up to 100 per cent)

necessarily reflect an optimal allocation. Thus, convergence to the European average should not be seen as a goal or a desirable outcome. One can observe a positive relation between both variables. Countries with higher levels of health spending (relative to the European average) tend to have higher levels of public health expenditures. There are two exceptions to this behaviour. The upper left quadrant displays one of those: countries where public health



Source: Eurostat; 2010−18 (or closest years); own calculations.

Figure 13.3 Current and public health spending relative to OECD average



Source: Eurostat; own calculations.

Figure 13.4 Growth of current and public health spending (2010−18 or closest year; left axis) and share of public health spending on current health spending (2018 or closest year; right axis)

spending lies above the European average, while total health spending is still lower than the average. This is the case of Portugal, Spain and Italy. The bottom right quadrant displays the second exception: the case where total health spending is above average, but public spending lies below it. Belgium, France, Luxembourg, the Netherlands and Germany are included in this group.

The following figure displays growth rates for total and public health spending since 2010, and plots the current share of public health spending in total health spending. Except for Greece, total health spending has increased over this period. Moreover, public health spending and total health spending have a tandem behaviour. However, the share of public health spending varies substantially across countries – reflecting different health systems' organization. Among the 30 different countries, ten have a share of public health spending lower than 10 per cent, while seven have a share higher than 70 per cent.

The growth of health spending reflects, among other factors, the roll-out of more expensive innovations, general economic growth (leading to higher aspirations and expectations for health care), as well as changes in demographic structures. In the opposite direction, this growth is also slowed by public policies for cost containment. However, in a context of low economic growth and fragile public finances, such growth imposes significant challenges to the fiscal sustainability of such spending. In fact, the total public health expenditure to GDP ratio is projected to more than double between 2015 and 2060, to 13.9 per cent in OECD countries. In an even more optimistic scenario, where cost-containment measures are enforced, the ratio would still increase by more than half, to reach 9.5 per cent (Oliveira Martins & de la Maisonneuve, 2015). Without significant room to increase revenues, governments need to balance the needs of different sectors and, in that sense, public health spending competes with other sources of public spending (such as education or social security for instance).

The health sector typically has restrictions to market competition as well as entry barriers. Firms in the health sector can also have different objective functions from traditional firms. Ethical behaviour, combined with regulation, may prevent purely profit-maximizing approaches. Additionally, the existence of public institutions or not-for-profit private institutions can also change market interactions between all agents. This is also a sector subject to significant uncertainty: both on the demand side (uncertainty with respect to disease incidence and severity), and on the supply side (uncertainty regarding the proper treatment and on its results). Such uncertainty, for a given patient at a certain moment of time, can result in abrupt changes on the health status and income. This is the cornerstone for protection schemes such as health insurance or public health care provision.

Unregulated private health insurance protects from uncertainty. However, such an instrument is likely to exclude high-risk patients through very high premiums. Traditional health insurance, in the presence of asymmetric information, is also likely to face the typical problems of adverse selection, moral hazards and other market failures. The existence of public intervention helps to counteract some of these issues and, particularly, contributes to ensuring protection for the entire population – preventing groups of patients from being excluded from the health system. Public intervention can take several forms, such as open enrolment rules, mandatory social health insurance or the direct provision of health services.

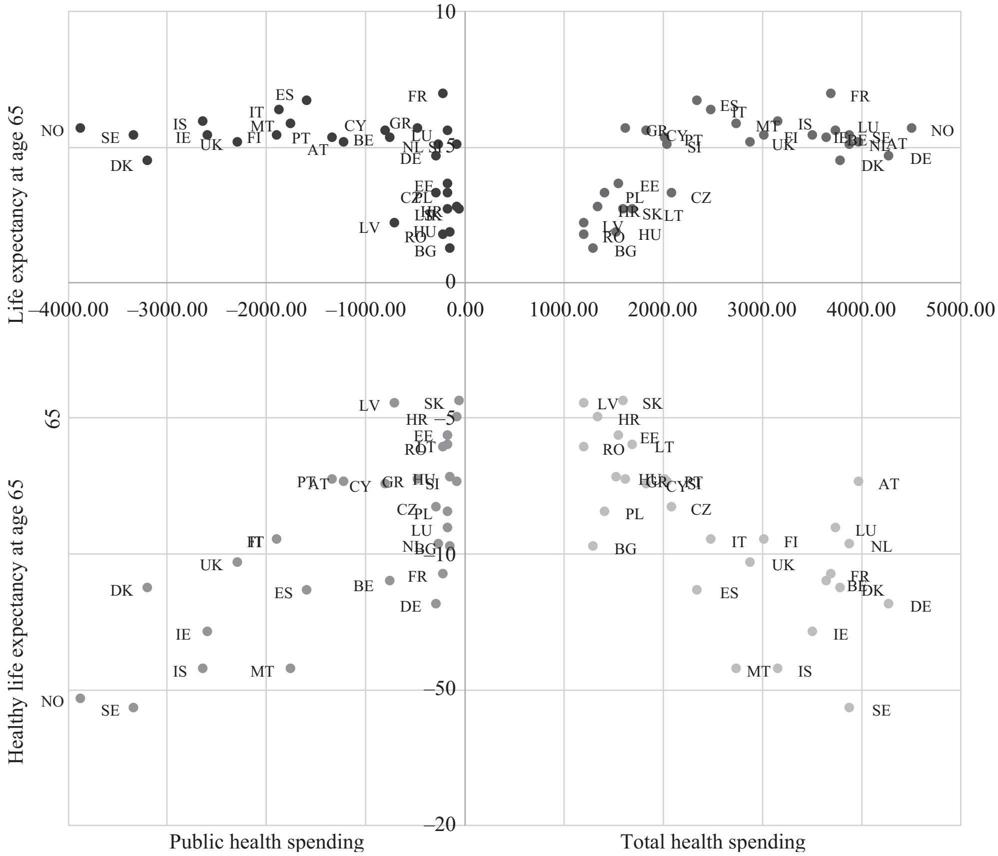
Different countries have organized such protection in different ways. This introduces an additional characteristic of the health system: patients do not usually pay directly to providers. In fact, the health sector has typically a third party, such as an insurance company or a public agency, which is responsible for collecting premiums or taxes from patients and making the payments for providers. With the exception of user fees and co-payments, a patient's payment occurs through such a third-party agent. Such an intermediary between consumers and producers of health care ensures protection and, at the same time, separates funding from provision. In fact, both funding and provision can be either private or public. Different health systems will have different combinations of such mechanisms. No evidence supports superiority of one system to the detriment of another, as different systems are the result of the evolution of a country's social protection schemes.

2. MEASURING PUBLIC HEALTH SECTOR EFFICIENCY

Public health spending has multiple objectives. Its main objective is to improve population health, which can be decomposed in terms of its quantity (longevity or life expectancy) and its quality (life quality indicators). This section will focus on both these objectives. Additionally, there are equity objectives of public health spending. In fact, such spending is used to promote access to health care, the protection of vulnerable populations (or specific population groups), as well as to ensure equity in the way the health system is financed. This section also accounts for the analysis of equity in health care access. However, the scope of equity in health system financing is not included here – as it is more related to tax system fairness and efficacy.

2.1 What is Public Health Sector Efficiency?

Health spending has increased over time and has resulted in improved health outcomes, such as longer and better lives. However, there is significant variability across countries, suggesting different degrees of efficiency of such spending. Figure 13.5 represents such a relation. The upper right quadrant represents the relation between life expectancy at age 65[3] and total health spending, while the upper left quadrant displays the relationship with public health spending. By the same token, the bottom right quadrant represents the relationship between total health spending and healthy life expectancy at age 65.[4] The bottom left quadrant displays the relation between that variable and public health spending.



Source: Eurostat; own calculations.

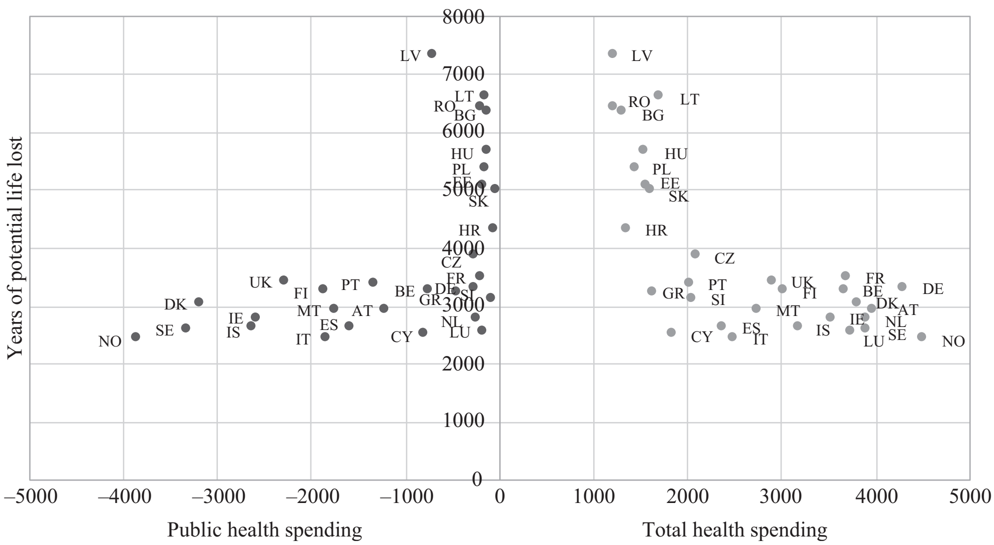
Figure 13.5 Relation between health expenditure and gains in life quality and quantity. Life expectancy at age 65 (top axis – 15 to 25 years) and healthy life expectancy at age 65 (bottom axis – 0 to 20 years), per capita public health spending (left axis – 0 to 4000 euros) and per capita total health spending (right axis – 0 to 5000 euros). (2018; Purchasing Power Standard per inhabitant)

There is a positive relation of both total and public per capita health spending to life quantity and quality (captured by life expectancy at age 65 and healthy life expectancy at age 65, respectively). However, one can notice significant variability in per capita health spending levels.

If we focus on the upper right quadrant, we can find examples of those variations. Taking the Czech Republic as an example, one can find other countries with similar levels of per capita total health spending, but with higher life expectancies at age 65. For instance, despite having similar spending levels, Slovenia's life expectancy at age 65 is higher by almost two years. Not only is Slovenia able to achieve higher life expectancies at age 65, but it is also able to achieve a healthy life expectancy at age 65 higher by roughly one year.

The same differences can be found when looking towards the relationship between health spending (public and total) with the years of potential life lost[5] across countries. Figure 13.6 shows that countries with higher levels of public and total health spending achieve, on average, lower levels of years of potential life lost. However, one can notice again significant differences in outcomes across countries with similar levels of health spending. The existence of such differences raises concerns about whether different health systems have different efficiency levels. If such an efficiency differential is true, then one could implement measures that, for the same health spending level, would improve health outcomes.

Economists define efficiency as the relation between inputs and outputs. A higher efficiency exists when the same level of inputs achieves higher outputs. Such a definition is built upon the concept of production functions, where, given a certain technology, inputs are combined and



Source: Eurostat; own calculations.

Figure 13.6 Relation between health expenditure and years of potential life lost (rate expressed per 100 000 age-standardized population under 70). Per capita public health spending (left axis – 0 to 5000 euros) and per capita total health spending (right axis – 0 to 5000 euros) (2017 or closest year; Purchasing Power Standard per inhabitant) processed into an output. In the health sector, the efficiency discussion can be easily translated to value-for-money or bang-for-buck analysis, where spending (public, private or both) is seen as an input in the process of producing health. Inefficiencies in the health sector occur when resources are not being used efficiently. Examples of inefficiencies in the health sector setting can be thought of as excessive hospital length of stay, over-prescribing, over-staffing, and wastage of stock among others.

Nonetheless, in economics, the concept of efficiency can mean different things. An important distinction needs to be made between technical and allocative efficiency. Technical efficiency implies that a certain unit, for instance a hospital, is producing at its maximum with the set of inputs it has (doctors, nurses, beds and so on). This implies that such a hospital is working on its production function, without any slack or waste of resources.

The concept of allocative efficiency is stricter, and it implies that money is not being spent on the wrong inputs. This allows for the possibility of selecting the optimal amount of inputs – instead of working with a fixed input level – such that a certain health care production level is achieved at the minimum possible cost. Therefore, from the set of technical efficient allocations (all production points lying on the production function), imposing an allocative efficiency requirement implies selecting the production function point compatible with the lowest cost. Figure 13.7 illustrates this idea.

The horizontal axis represents health spending, which can be thought of as the key input used in the production of health. The vertical axis represents such health – measured in both quality and quantity of life terms, which can be interpreted as the quality adjusted life years (QALY). As seen in Figure 13.7, countries with different levels of health spending can have different outcomes. The line in the plot represents the production possibility frontier, meaning that all points on and below the line can be attained by the health system. However, given a certain level of health spending, no point above the production possibility frontier can be achieved.

A health care system represented by point A is below the production possibility frontier. This means that the current spending level is not achieving the maximum possible output. Point A is

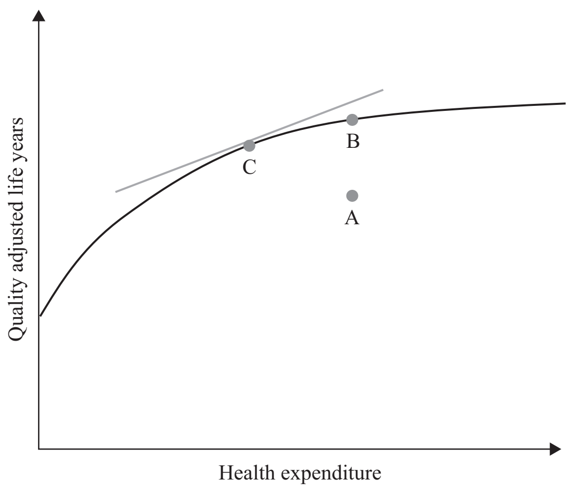


Figure 13.7 Illustration of technical and allocative efficient bundles

therefore technically inefficient. If efficiency inducing measures were to be implemented, the country health system could eventually evolve up to point B (the example of the Czech Republic versus Slovenia mentioned above).

However, while point B is technically efficient, it is still allocatively inefficient. Cost-effectiveness thresholds ensure that only treatments that yield QALYs at a reasonable cost are adopted. If we assumed this to be the case, and we consider the orange line in the plot to represent the 'reasonable' threshold, then point B is allocatively inefficient. Although health spending is being used in an efficient way, we are spending an 'unreasonable' amount of money for a given output. An alternative would be to move along the production possibility frontier towards point C. This point would be both technically and allocatively efficient. Point C would therefore be preferred to B since the reduction in health spending involved in getting from C to B would compensate for losses in QALYs. Note that this reasoning hinges on the assumption that we are able to define the 'reasonable threshold' for the cost on an additional quality adjusted life year – which is one of the main challenges in cost-effectiveness analysis.

Hence, the concept of efficiency can be represented by a generic indicator such as the ratio of resources to output produced. Under this setting, any deviation from the production possibility frontier would represent inefficiency. However, despite the theoretical framework under which efficiency is analysed seeming straightforward, practical difficulties in applying such a concept to the health system and, particularly, to the public health sector remain.

One of the main challenges is the determination of the appropriate inputs that should be considered. Using different inputs will result in different estimates which may bias conclusions. Additionally, one needs to decide on the disaggregation level of inputs. At one extreme, a single measure of aggregate inputs might be used – such as the health spending example described above. This approach assumes that agents can choose freely their inputs, given such a spending cap. However, for some analyses, such an assumption is unrealistic. For instance, in a short-run efficiency analysis for hospitals, it may be better to specify inputs at a more disaggregate level – such as physicians, nurses, beds, operating rooms and so on.

A second challenge is the relation between inputs and outputs. In Figure 13.4, we have seen that, for a similar public health spending level, Slovenia achieved higher life expectancies than the Czech Republic. A crude efficiency analysis would then conclude that the Czech Republic has an inefficient health system. However, there are factors other than health resources that can contribute for better health outputs. Education, income or biological characteristics can affect health outcomes. Additionally, many other factors can explain health outcomes, such as health behaviours (smoking, alcohol consumption, drug use, obesity and so on). Such multiplicity of factors prevents the development of robust measures of comparative efficiency, as noted by McDaid et al. (2012). Moreover, Cylus et al. (2017) argue that health care is tailor made, with consequent variations in clinical needs, social circumstances and personal preferences – affecting how inputs are consumed and outputs are produced.

A third challenge relates to the output choice. The relation between inputs, intermediate outputs and health outcomes is complex and multifaceted (Medeiros & Schwierz, 2015). As an example, hospital discharges are usually seen as outputs when in fact they are an intermediate output: health care activities do not necessarily have an immediate impact on improving health. Inputs and outputs differ in often inadequately measured dimensions such as in quantity and quality, or volume and value. Outputs tend to represent health outcomes, such as life expectancy, healthy life expectancy or mortality rates. Depending on the level at which the analysis is being made, outputs may also represent volume of production, such as number of surgeries or doctor appointments at a given hospital. In fact, research is usually constrained to examining efficiency based on outputs instead of on outcomes. Such measures are manifestly inadequate, as they fail to capture variations in the effectiveness and value of the health care provided. For micro-based analysis, progress is being made in the use of common international metrics, which will allow more solid comparisons. Patient-reported outcome measures (PROMs) and EQ-5D or SF-36 questionnaires are just some examples.

Finally, as highlighted by McDaid et al. (2012), a production functions approach typically does not account for other goals of the health system such as user satisfaction, equity or financial protection. When analysing public health spending one should also consider equity and unmet needs as relevant outputs. This is a significant limitation of current research.

The concept of a production function is broad and flexible. Such a framework can be used to assess very detailed micro-units (such as a physician's office or hospitals) or macro-units (such as the entire health system and cross-country comparisons). Even if all factors mentioned above are accounted for, an important question remains: do all countries or units have the same production function? If some institutional features of some units or health systems prevent them from producing as much health as other systems, should this be considered as inefficiency? Can it be the case that some identified inefficiencies result from the fact that the actual production functions are not the same across units or countries?

Efficiency analyses are often made using descriptive statistics, with benchmark and peer group analysis. However, econometric models can also be used to perform a more robust analysis, controlling for some of the issues identified above. The most common methods are Stochastic Frontier Analysis (Aigner et al., 1977; Meeusen & van den Broeck, 1977) and Data Envelopment Analysis (Banker et al., 1984; Boussofiane et al., 1991; Charnes et al., 1978).

Stochastic frontier analysis, whose methods have been extensively developed, among others by Kumbhakar and Lovell (2000), allows us to estimate the production frontier with an error term that has two components – a random error and a strictly non-negative term, which captures inefficiency. Data envelopment analysis is a non-parametric method for the estimation of the production frontier using linear programming techniques – without assuming a particular functional form for the frontier. Jacobs et al. (2006) provide a set of applications of both these methods in the health sector.

Next, we provide some brief examples on the main methodologies that have been followed to assess efficiency in the health sector. We distinguish such studies between macro-analysis – usually focused on comparing countries and health systems – and micro-analysis – focused typically on a set of individual structures, for instance hospitals.

2.2 A Macro Perspective: Comparing Health System Efficiency

In the empirical literature, a large number of efficiency analyses have been published with the goal of providing rankings for health systems. However, the problems highlighted above prevent fair comparisons being made. Despite significant contributions, empirical studies still find significant challenges.

The World Health Report (2000) presents an extended efficiency analysis for over 190 national health systems, based on an empirical estimate of the production function. This production function includes indicators for: overall health level, distribution of health, overall level of responsiveness, distribution of responsiveness and distribution of financial contributions. By including these five dimensions, the WHO report clearly recognizes goals for a health system other than health outcomes. According to these estimates, France was considered to have the most efficient health system − achieving 99.4 per cent of its potential. Hollingsworth and Wildman (2003) used alternative parametric and non-parametric methods to re-estimate the WHO data. They found that different methods yielded different results. Using the same dataset, Green (2004) suggests that there was considerable heterogeneity masquerading as inefficiency.

Joumard et al. (2010) argue that efficiency is correlated with quality of care, even though the existing quality indicators still do not have sufficiently wide coverage to make solid cross-country comparisons. Using DEA, they estimate that life expectancy at birth could be raised by more than two years, holding health care spending constant. OECD (2014) points to significant inefficiencies in the use of resources based on major geographical variations in medical practices across and within countries. Medeiros and Schwierz (2015) estimate EU health care system efficiency using different models, with different combinations of outputs and inputs – including controls for environmental factors. Their results are aligned with previous empirical research, suggesting that EU life expectancy at birth could be increased by 1.8 years.

Health system comparisons often ignore different public and private health spending mixes. However, such a mix is far from being consensual. A common discussion is whether private spending achieves higher efficiency levels than public spending. On the one hand, one can argue that private spending has stronger incentives for efficiency. On the other hand, such private spending does not account for market failures and can lead to a mismatch between profit maximizing and public health goals. Existing literature on efficiency between private and public provision of health care services shows inconclusive evidence: one cannot generalize which ownership model is best across countries or even within countries over time (Hsu, 2010). Still, The European Commission and the EU's Economic Policy Committee (2010), and the European Commission (2014) have suggested that inefficiencies in health care can be associated with a non-optimal mix between private and public funding.

Hsu (2010) summarizes the main evidence regarding public and private health spending. Hollingsworth (2008) conducted a meta-analysis of 317 published works on efficiency measures and concludes that 'public provision may be potentially more efficient than private'. However, such a conclusion largely depends on the specific country and health system under analysis. For instance, Lee et al. (2009) determined that non-profit hospitals in the United States were more efficient than for-profit hospitals. In Taiwan, Chang et al. (2004) found the private sector to be more efficient than the public sector. In Germany, evidence is mixed, with authors finding that private hospitals are less efficient than public hospitals (Helmig & Lapsley, 2001), while others concluded the inverse or found no difference (Staat, 2006). In Switzerland, hospitals' efficiency levels were not predisposed towards inefficiency by type of ownership (Steinnmann & Zweifel, 2003).

2.3 A Micro Perspective

Multiple studies have been developed at a micro-level, studying the relation between particular inputs in order to find optimal combinations which could foster efficiency. If the previous macro-studies are usually more focused on the concept of technical efficiency, the micro-based studies often have information on input costs, which can be used to analyse allocative efficiency.

Jensen and Morrisey (1986a, 1986b) explore input substitution effects using flexible functional form production functions. They find strong substitution effects between nurses and doctors. Such results were reinforced by the literature. Brown (1988) shows that office-based physicians under-utilize their assistants. Similar effects are identified in studies conducted on dental auxiliaries (Liang & Ogur, 1987), mental health professionals (Okunade & Murthy, 2008) and even pharmacists (Okunade & Suraratdecha, 1998). Estimates suggest that one physician extender, for instance nurses or physicians' aides, could substitute 25 per cent to 50 per cent of a doctor's services.

Other micro-based studies focus on assessing the relative efficiency of different hospitals within a certain country. There are unresolved debates about the magnitude of economies of scale and economies of scope in health services, particularly in the hospital sector – as described by Giancotti et al. (2017). Discussions are common regarding the existence of gains with vertical integration or relative to the optimal size of a hospital.

The introduction of Diagnosis-Related Groups (DRG), which allow us to classify and segment patient groups, have been an important step towards improving the reliability of health system comparison results. Such patient classification systems allow for a more detailed analysis, controlling for case-mix differences across hospitals. Cylus et al. (2016) provide examples that show that DRG systems can be used to assess hospital efficiency at different levels: at the macro-level to compare total hospital costs per case, while adjusting for case mix; at the meso-level to compare actual costs per DRG across different hospitals; and at the micro-level to compare a particular hospital cost structure against average costs of treatment in other hospitals.

3. AN EFFICIENCY ANALYSIS FOR THE PUBLIC HEALTH SECTOR

This section illustrates some of the trade-offs mentioned above by conducting an efficiency analysis of public health spending for the same set of European countries. Variables were collected from the Health database from Eurostat and the World Health Organization (Eurostat, 2020; WHO, 2020), with a full description available in the annex. Table 13.1 describes the main descriptive statistics for the variables included in the analysis.

We pursue two alternative analyses that complement each other. On the one hand, we estimate a cost function using longitudinal data. On the other hand, we estimate production functions using information for 2018. These two approaches allow us to deal with the limited and incomplete data challenge.

The production function analysis, as described above, focuses on explaining the production of a specific output, with a given set of inputs. The cost function estimation is an approach that looks to the same problem from a different perspective: it estimates the contribution of output (which is a function of inputs) and input prices to an expenditure variable.

As discussed before, one needs to analyse the impact of health spending on different outcomes. Five different outputs are considered in the 2018 cross-sectional analysis. However, due to data availability issues, only three outputs will be used for the panel estimation. The first outcome is life expectancy at age 65 to capture the goal of promoting longevity. Health spending also has the goal of promoting higher life quality. For that reason, the second output is healthy life expectancy at age 65. The third output is the potential years of life lost as an indicator of premature deaths. Finally, the last two outputs capture equity goals of health spending: unmet needs and catastrophic health expenditures.

Table 13.1 Key variables descriptive statistics (average across countries for 2018 or closest year)

Variable	Units	Mean	Standard Deviation	Minimum	Maximum
Life expectancy at 65	Years	19.48	1.56	16.20	21.90
Healthy life expectancy at 65	Years	9.32	3.10	4.40	15.70
Unmet health care needs	% of population	25.26	9.55	9.40	41.80
Potential years of life lost	per 100k pop (<70yr)	3,820	1,419	2,425	7,329
Catastrophic health spending	% of population	7.96	5.27	1.42	18.38
Public health spending	PPS per inhabitant	1076.94	1124.27	38.89	3852.59
Other health spending	PPS per inhabitant	1536.36	1025.61	517.47	4.028,6
Government balance	% of GDP	−0.25	1.65	−3.70	3.10
Risk of poverty	% of population	21.37	5.53	12.20	32.80
Low education levels	% of population	23.78	9.14	11.70	49.80
Obesity	% of population	16.31	3.23	9.10	25.20
Low fruit and vegetable intake	% of population	35.87	10.64	16.10	65.10
Daily smokers	% of population	19.64	4.94	9.80	28.20

Unmet needs represent the share of the population unable to access the health system due to waiting times, financial constraints or distance/transportation restrictions. On the other hand, catastrophic health expenditures represent the share of the population that bears large health expenditures (greater than 10 per cent of total household income). Catastrophic expenditures can be seen as a proxy for the value of health protection, as the role of health insurance (private or public) is to protect individuals from negative health shocks (and the respective financial burden). The goal of the health system should be to attain high levels of life expectancy and healthy life years, while minimizing potential years of life lost, unmet needs and catastrophic expenditures.

To estimate the cost function, we pursue a fixed-effects panel approach using data from 2008 to 2018 for a set of 28 countries. A stochastic frontier model estimation was used, with time-invariant inefficiency. On this estimation, the dependent variable is health expenditure. We consider three different models. The first uses total health expenditure as a dependent variable. The second model uses only public health expenditure (direct government financing schemes). The third model uses the remaining health expenditure.

The independent variables on the cost function are the relevant outputs, input prices and other control variables. In this analysis we have included three out of the five outputs described above – due to data availability constraints. We assume that input prices have not changed significantly over time. Thus, these were not included in the estimation since their effect is captured by the fixed-effects term. The same holds for the set of time-invariant characteristics that could affect the health system cost (such as the type of health system or population charac-

Table 13.2 Stochastic frontier estimates for cost functions (panel data analysis for 2008−18 data)

	(1) Total health expenditure	(2) Public health expenditure	(3) Other health expenditure
Life expectancy at age 65	0.5582	−1.0536	0.1367
	(0.4502)	(1.2707)	(0.4774)
Healthy life expectancy at age 65	−0.0080	−0.1822	−0.0030
	(0.0512)	(0.1446)	(0.0548)
Potential Years of Life Lost	−0.7183***	−1.1052***	−0.6496***
	(0.1028)	(0.2909)	(0.1113)
Share of OOP on lagged total health expenditure	0.1354	0.3302	0.5949***
	(0.1183)	(0.3125)	(0.1100)

Notes: *,**, *** indicate significance at 10 per cent, 5 per cent and 1 per cent level, respectively. Standard errors in parentheses. Variables in logarithms. Twenty-eight countries included. Time-invariant inefficiency.

teristics). Additionally, we introduce the share of out-of-pocket payments in lagged total health expenditure as a proxy for financial protection. Table 13.2 shows our main model estimates.

Results suggest that changes in longevity or life quality do not significantly affect health expenditure (neither public nor private). However, there is a significant impact from the Potential Years of Life Lost on all types of health spending. In fact, estimates suggest that lower levels of premature mortality are associated with higher levels of health spending. Such an effect is particularly strong in public health expenditure compared to other health spending. These results reinforce the importance of health systems in reducing premature mortality.

A cross-sectional production function approach was also performed to complement the previous analysis. This approach uses 2018 data (or closest available year) for the same set of 28 countries. Full estimation results are available in the Appendix.

This estimation was done by fitting a stochastic frontier model with a normal/half-normal distribution for each of the five different outcomes. In this production function approach the dependent variable is each outcome.

Independent variables represent inputs that contribute to the production of such an output. In fact, this set of outcomes can be explained by multiple factors. We are particularly interested in analysing the role of health spending – which will be divided into four groups of health spending. We are considering health spending as monetary inputs to the production of each outcome considered.

Because different health systems are organized in different ways, interaction dummy variables are introduced to capture the role of public health spending in different health systems and different organizations. We classify health systems into four different groups[6] according to the OECD classification and Health Systems in Transition information (Bohm et al., 2012).

The government balance is also introduced as an indicator for fiscal space. Large government deficits might prevent significant increases in public health spending, while government surplus might allow such an increase. Additionally, a set of control variables is included as they are likely to affect outcomes. If two countries have the same health system, but the population of one country has a high smoking prevalence, one should expect life expectancy to be lower in that country (Fuchs, 1974). Some controls have been dropped due to data availability and collinearity issues.[7] The final model specification includes the proportion of people at risk of poverty, with low education levels, with obesity, without daily fruit or vegetable consumption and the share of daily smokers in the population.

The inclusion and availability of additional variables is an advantage relative to the cost estimation. However, results should be taken with a grain of salt since this cross-sectional analysis has a very low number of observations. This prevents very precise estimates. Still, some conclusions can be derived.

Results suggest that public health spending is associated with improvements in longevity and life quality, and with declining premature mortality. Interestingly, countries with National Health Services are associated with significant and large reductions in catastrophic expenditures and unmet needs. This suggests that these systems (usually financed through public health spending) have an important impact in terms of promoting access and providing financial protection for negative health shocks.

It is interesting to notice the role of voluntary insurance in countries with a National Health Service. Results suggest that such voluntary financing schemes play a role in improving longevity and reducing premature mortality. However, they reinforce unmet needs and catastrophic expenditures.

We also find that the inclusion of the government balance as an explanatory factor is only relevant for countries with a National Health Service. For these countries, higher surpluses (which may signal tighter control of public spending) are associated with higher levels of unmet needs and catastrophic expenditures.

Socio-economic factors also play a role in these outcomes. Higher levels of poverty in the population are associated with higher levels of potential years of life lost, catastrophic health expenditures and unmet needs. Moreover, less educated individuals are associated with higher levels of potential years of life lost and unmet needs.

Health behaviour variables are also usually aligned with common intuition. Low consumption of fruits and vegetables impacts life expectancy negatively, while increasing potential years of life lost. Nonetheless, positive effects are found on unmet needs and catastrophic health expenditures. Populations with more obese individuals are also likely to have higher catastrophic expenditures, and lower levels of unmet needs.

Besides analysing the relation between inputs and outputs in the production functions, stochastic frontier models allow us to comment on the relative efficiency of different countries. The model predicts the distance from each country relative to the efficient benchmark (the production frontier). Our cross-sectional model has several limitations – mainly related with data availability and a small number of observations. Nonetheless, one can derive some conclusions relative to inefficiency.

Table 13.3 displays average inefficiency scores for all countries for each of the five different models estimated. Overall, estimates suggest relatively small inefficiencies. On average, countries are close to their production frontier. However, models for unmet health care needs and catastrophic health spending display higher inefficiency scores relative to the remaining models. Such a pattern suggests inefficiencies in the way health systems deliver financial protection and ensure access to their populations. Thus, access and financial constraints seem to be more relevant than health outcomes to explain health system inefficiency.

Table 13.3 Average technical inefficiency estimates (stochastic frontier for production functions: cross-sectional 2018 data)

Model (dependent variable)	Average technical inefficiency
Life expectancy at 65	3%
Healthy life expectancy at 65	5%
Unmet health care needs	10%
Potential years of life lost	6%
Catastrophic health spending	12%

Different countries have different efficiency scores for each variable, with no systematic pattern displayed. However, we can identify countries which, on average, have higher inefficiency scores for these five outputs. According to this model, the top five countries displaying higher inefficiencies are Latvia, Greece, Netherlands, Slovakia and Slovenia. Similarly, it is possible to identify countries with low inefficiency scores. The top five countries with the most efficient health systems are Belgium, Bulgaria, Poland, Hungary and Romania.

Interestingly, no particular health system is associated with higher or lower inefficiencies. In fact, both groups of countries (more and less efficient) have Eastern European countries and similar health systems. With few exceptions, all countries in both groups have health systems financed on a societal basis with private or societal provision. Thus, no specific pattern is associated with having higher or lower inefficiency scores. Results suggest that different health systems can be equally efficient in the production of health. Drivers for efficiency might be related to specific mechanisms within each health system, which are not easily identified or observed on these macro comparisons.

4. CONCLUSIONS

Health spending has increased over time in developed economies. In the context of low levels of economic growth and fragile public finances, there are concerns regarding the growth limit for public health spending. Therefore, the quest for efficiency in the health sector remains a major priority for governments.

However, efficiency in the health sector, and particularly within public spending, is very hard to define. In fact, the health system has multiple goals. Additionally, different types of health spending contribute differently to attaining each of those goals.

Research on health spending efficiency has been discussed, both from macro- and microeconomic perspectives. Efficiency analysis is often conducted within a production function approach. Still, there are questions that remain unanswered: for instance, researchers struggle to determine which inputs and outputs to analyse. Different choices lead to completely different conclusions. Additionally, production function approaches typically do not account for other goals of the health system such as user satisfaction, equity or financial protection.

In this chapter we perform an efficiency analysis using both cost and production function estimations. With our estimations we highlight that health spending sources contribute differently to the numerous health system goals. Results suggest that public health spending has a role in promoting access and providing financial protection to the population. Nonetheless, other forms of health spending also contribute to the general goals of the health system.

When defining measures to improve efficiency of public health spending, governments should not ignore the contribution of such expenditure to attaining some of the goals of the health system. This seems particularly relevant in National Health Service settings.

Further research is required to better understand the connection between the types of spending and the organization of health systems. Also, there are complement and substitution effects between the different sources of health spending which might be interesting to explore.

APPENDIX

Table 13.4 Variables description

Variable	Definition	Units
Life expectancy at 65	Life expectancy at age 65	Years
Healthy life expectancy at 65	Healthy life expectancy at age 65	Years
Potential years of life lost	Per 100 000 age-standardized population under age 70	Years per 100 000
Unmet health care needs	Proportion of self-reported unmet needs for health care (Financial reasons, distance or transportation, waiting list)	%
Catastrophic health expenditures	Population with household expenditures on health greater than 10% of total household income	%
Public health spending	Health expenditure from compulsory government financing schemes	PPS per inhabitant
Other health spending	Total current health expenditure net of health expenditure from compulsory government financing schemes	PPS per inhabitant
Poverty	Proportion of population at risk of poverty or social exclusion	%
Low education levels	Proportion of people with education level lower than primary and lower secondary education (levels 0−2)	%
Obesity	Proportion of obese people according to reported BMI	%
Low fruit & vegetable intake	Proportion of people that do not consume daily fruit and vegetables	%
Daily smokers	Proportion of daily smokers in the population	%

Table 13.5 Stochastic frontier estimates for production function (2018 or closest available year)

	(1) Life expectancy at age 65	(2) Healthy life expectancy at age 65	(3) Potential years of life lost	(4) Unmet health care needs	(5) Catastrophic Health Expenditures
Compulsory public insurance	0.0544**	0.1579*	−0.0995*	0.4501***	0.4564***
	(0.0228)	(0.0816)	(0.0586)	(0.0777)	(0.0790)
Compulsory private insurance	0.0348*	−0.0741	−0.0299	0.2278***	0.3857***
	(0.0188)	(0.0681)	(0.0489)	(1.3377)	(0.0870)
Voluntary insurance x NHS	0.7965**	−0.9626	−2.4816***	4.5521***	6.5758***
	(0.3309)	(1.2576)	(0.9320)	(1.3377)	(1.2516)
Voluntary insurance x Other	0.0137	0.2255***	0.0133	−0.1528**	−0.8459***
	(0.0200)	(0.0761)	(0.0549)	(0.2683)	(0.1325)
Out-of-pocket payments	−0.1951**	−0.0656	0.2730	−1.5243***	−0.1821*
	(0.0847)	(0.2753)	(0.1960)	(0.2683)	(0.1008)
Government Balance x NHS	0.0893*	−0.1690	−0.0264	0.9212***	0.9749***
	(0.0512)	(0.1803)	(0.1301)	(0.1855)	(0.1064)
Government Balance x other systems	0.0042	−0.0134	−0.0880	0.0020	−0.0238
	(0.0087)	(0.0275)	(0.0181)	(0.0227)	(0.0294)
NHS	−3.874**	4.4576	13.3709***	−22.3682***	−31.9993***
	(1.6544)	(6.3396)	(4.7031)	(6.6893)	(6.5528)
NHI	−0.2847***	−0.3735	0.1623	−1.7238***	−0.8132***
	(0.0980)	(0.3198)	(0.2255)	(0.3183)	(0.2361)
SHI	0.0072	−0.2075	−0.1656	−0.0060	1.5613***
	(0.0574)	(0.1992)	(0.1350)	(0.2333)	(0.2228)
Public insurance x NHS	0.0002**	0.0005	−0.0008***	0.0007**	−0.0017***
	(0.0001)	(0.0003)	(0.0002)	(0.0003)	(0.0004)
Public insurance x NHI	0.0002**	−0.0001	−0.0002	0.0011***	0.0010***
	(0.0001)	(0.0003)	(0.0002)	(0.0003)	(0.0002)

(*Continued*)

Table 13.5 (Continued)

	(1) Life expectancy at age 65	(2) Healthy life expectancy at age 65	(3) Potential years of life lost	(4) Unmet health care needs	(5) Catastrophic Health Expenditures
Public insurance x SHI	0.0005	0.0019	−0.0002	0.0048***	−0.0084***
	(0.0004)	(0.0012)	(0.0008)	(0.0013)	(0.0009)
Poverty	0.0739	0.2316	0.5887***	1.2626***	1.6751***
	(0.0553)	(0.1885)	(0.1326)	(0.1637)	(0.3100)
Low education levels	−0.0020	−0.0689	−0.5430***	−0.2677**	0.4590
	(0.0557)	(0.1692)	(0.1107)	(0.1342)	(0.3064)
Obesity	−0.3225**	−0.3456	0.3000	−0.6879**	0.8928***
	(0.1407)	(0.3941)	(0.2546)	(0.3320)	(0.2259)
Low fruit & vegetable intake	−0.1651**	0.2928	0.3289**	−1.2307***	−1.0323***
	(0.0649)	(0.2023)	(0.1383)	(0.2000)	(0.1820)
Tobacco consumption	0.2664**	0.0490	−0.2817	1.3607***	0.1235
	(0.1299)	(0.4415)	(0.3204)	(0.4781)	(0.1820)

Notes: *,**, *** indicate significance at 10 per cent, 5 per cent and 1 per cent level, respectively. Standard errors in parentheses. Normal/Half-normal distribution. Variables in logarithms. Twenty-eight countries included in models (1) – (3). Twenty six included in models (4) – (5) due to data availability. NHS – National Health Service, NHI – National Health Insurance, SHI – Social Health Insurance (Baseline group: Estatic Social Health Insurance.

NOTES

1. Countries included: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom.
2. Countries included: Australia, Austria, Belgium, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom, United States.
3. Life expectancy at age 65 is preferred to life expectancy at birth as it is an indicator more related with the health system.
4. Number of remaining years that a person at age 65 is expected to live without any health problems.
5. Measures premature mortality, giving more weight to deaths among younger people. It estimates the average years a person would have lived if she/he had not died prematurely (defined as before age 70).
6. Systems financed and regulated by the State: *National Health Service* (state provision), *National Health Insurance* (private provision), Systems financed on a Societal basis: *Social Health Insurance* (private provision) and *Estatic Health Insurance* (private or societal provision − baseline group against which the other system will be compared).
7. Full details can be provided by the authors upon request.

REFERENCES

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*(1), 21–37.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*(9), 1078–1092.

Bohm, K., Schmid, A., Gotze, R., Landwehr, C., & Rothgang, H. (2012). Classifying OECD healthcare systems: A deductive approach. *TranState Working Papers*, *165*.

Boussofiane, A., Dyson, R. G., & Thanassoulis, E. (1991). Applied data envelopment analysis. *European Journal of Operational Research*, *52*(1), 1–15.

Brown, D. M. (1988). Do Physicians Underutilize Aides? *The Journal of Human Resources*, *23*(3), 342. https://doi.org/10.2307/145833

Chang, H., Cheng, M., & Das, S. (2004). Hospital ownership and operating efficiency: Evidence from Taiwan. *European Journal of Operational Research*, *159*, 513–527.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.

Cylus, J., Papanicolas, I., & Smith, P. C. (Eds). (2016). *Health system efficiency: How to make measurement matter for policy and management*. WHO Regional Office for Europe.

Cylus, J., Papanicolas, I., & Smith, P. C. (2017). How to make sense of health system efficiency comparisons? *European Observatory on Health Systems and Policies Policy Briefs*, *27*, 28.

European Commission (2014). Identifying fiscal sustainability challenges in the areas of pension, health care and long-term care policies. *European Economy - Occasional Papers*, *201*. https://doi.org/10.2765/85081

European Commission, & Economic Policy Committee (2010). Joint report on health systems. *European Economy - Occasional Papers*, *74*.

Eurostat (2020). *Eurostat Health Statistics Database*.

Fuchs, V. R. (1974). *Who Shall Live?: Health, Economics and Social Choice*. World Scientific.

Giancotti, M., Guglielmo, A., & Mauro, M. (2017). Efficiency and optimal size of hospitals: Results of a systematic search. *PLOS ONE*, *12*(3), e0174533. https://doi.org/10.1371/journal.pone.0174533

Green, W. (2004). Distinguishing between heterogeneity and inefficiency: Stochastic frontier analysis of the World Health Organization's panel data on national health care systems. *Health Economics*, *13*, 959–980.

Helmig, B., & Lapsley, I. (2001). On the efficiency of public, welfare and private hospitals in Germany over time: A sectoral data envelopment analysis study. *Health Services Management Research*, *14*, 263–274.

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. *Health Economics*, *17*(10), 1107–1128. https://doi.org/10.1002/hec.1391

Hollingsworth, B., & Wildman, J. (2003). The efficiency of health production: Re-estimating the WHO panel data using parametric and non-parametric approaches to provide additional information. *Health Economics*, *12*(6), 493–504. https://doi.org/10.1002/hec.751.

Hsu, J. (2010). The relative efficiency of public and private service delivery. *World Health Report Background Paper*, *39*, 9.

Jacobs, R., Smith, P. C., & Street, A. (2006). *Measuring Efficiency in Health Care*. Cambridge University Press. https://doi.org/10.1017/CBO9780511617492

Jensen, G. A., & Morrisey, M. A. (1986a). The role of physicians in hospital production. *The Review of Economics and Statistics*, *68*(3), 432. https://doi.org/10.2307/1926020

Jensen, G. A., & Morrisey, M. A. (1986b). Medical staff specialty mix and hospital production. *Journal of Health Economics*, *5*(3), 253–276. https://doi.org/10.1016/0167-6296(86)90017-2

Joumard, I., André, C., & Nicq, C. (2010). Health care systems: Efficiency and institutions. *OECD Economics Department Working Papers*, *769*(769). https://doi.org/10.1787/5kmfp51f5f9t-en

Kumbhakar, S. C., & Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press.

Lee, K., Yang, S., & Choi, M. (2009). The association between hospital ownership and technical efficiency in a managed care environment. *Journal of Medical Systems*, *33*, 307–315.

Liang, N. J., & Ogur, J. D. (1987). *Restrictions on Dental Auxiliaries: An Economic Policy Analysis*. Federal Trade Commission.

McDaid, D., Merkur, S., Maresso, A., Kossarova, L., & Mossialos, E. (2012). *Efficiency and Health Care* (18 (3); EuroHealth, p. 40). European Observatory of Health Systems and Policies.

Medeiros, J., & Schwierz, C. (2015). Efficiency estimates of health care systems. *European Economy - Economic Papers*, *549*, 60.

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, *18*(2), 435–444.

OECD (2014). Geographic variations in health care: What do we know and what can be done to improve health system performance? *OECD Health Policy Studies*. http://dx.doi.org/10.1787/9789264216594-en

OECD (2019). *OECD Health Statistics 2019*. OECD.

Okunade, A. A., & Murthy, V. N. R. (2008). Are physician and non-physician providers of outpatient mental healthcare substitutes or complements? A conceptual clarification. *Health Care Management Science*, *11*(4), 393–398. https://doi.org/10.1007/s10729-007-9044-5

Okunade, A. A., & Suraratdecha, C. (1998). Cost efficiency, factor interchange, and technical progress in US specialized hospital pharmacies. *Health Economics*, *7*(4), 363–371. https://doi.org/10.1002/(SICI)1099-1050(199806)7:4<363::AID-HEC346>3.0.CO;2-Q

Oliveira Martins, J., & de la Maisonneuve, C. (2015). The future of health and long-term care spending. *OECD Journal: Economic Studies*, *2014*(1), 61–96. https://doi.org/10.1787/eco_studies-2014-5jz0v44s66nw

Staat, M. (2006). Efficiency of hospitals in Germany: A DEA-bootstrap approach. *Applied Economics*, *238*, 2255–2263.

Steinnmann, L., & Zweifel, P. (2003). On the (in)efficiency of Swiss hospitals. *Applied Economics*, *35*(3), 361–370. https://doi.org/10.1080/00036840210167183

WHO (2020). *Global Health Expenditure Database*.

World Health Organization (Ed.) (2000). *The World Health Report 2000: Health systems: improving performance*. WHO.

14. Evaluating the efficiency of social protection spending

David Coady and Samir Jahan[1]

1. INTRODUCTION

Social protection systems play a crucial role in protecting households from income shocks and poverty and in achieving distributional objectives. These systems include both social insurance and social assistance programs. Social insurance programs, such as unemployment benefits, protect workers from loss of income due to unemployment.[2] Social assistance transfers protect households from poverty or low income and tend to be categorical (such as universal child benefits) or means tested (such as Guaranteed Minimum Income programs). Evaluation of social protection spending needs to examine both whether spending levels are adequate and spending design efficient. Countries can fail to achieve social objectives either because spending is inadequate or it is inefficient, or some combination of both.

This chapter focuses primarily on the efficiency of social protection spending. As for any component of public spending, the efficiency of social protection spending needs to be evaluated in terms of some clearly specified social objectives that it is meant to achieve. In the context of social protection spending, these social objectives are typically expressed in terms of reducing income poverty or income inequality. Social protection systems are essentially a way of transferring income from higher-income groups to lower-income groups (the 'poor') to make the distribution of disposable incomes (that is, income after direct taxes and transfers) less unequal than the distribution of market incomes (that is, income before direct taxes and transfers). Empirical evaluations of the efficiency of social protection spending also need to allow for the fact that other factors can independently affect poverty and inequality, such as the overall level of national income, the degree of monopoly or monopsony power in product or factor markets, the distribution of education outcomes and other factors that affect the distribution of market incomes.

The primary objective of the chapter is to review the conceptual and empirical issues that need to be addressed when evaluating the efficiency of social protection spending—henceforth we simply refer to these as 'spending' and 'spending efficiency.' It is not intended to be a review of the empirical literature, which is vast and beyond the scope of the chapter.[3] Reflecting this, the empirical analysis presented is intended solely as illustrative of these conceptual and empirical issues. This analysis also provides a basis for determining how one should move from an empirical analysis of spending efficiency to discussing policy implications. The chapter also focuses on spending as opposed to the taxes used to finance it. While the tax side can also have important distributional implications, existing empirical analyses consistently find that the bulk of the impact on poverty and inequality comes from the spending side of the fiscal equation (Clements et al., 2015).

The format of the chapter is as follows. Section 2 sets out a conceptual framework for the analysis of spending efficiency. Section 3 identifies other costs policy makers need to consider when designing transfer schemes and which may require trade-offs with spending efficiency. Using publicly available data for European countries, Section 4 shows how the conceptual framework can be applied in practice to evaluate spending efficiency. Section 5 discusses how the results from empirical analysis can be used to inform policy choices for enhancing spending efficiency. Section 6 concludes.

2. CONCEPTUAL FRAMEWORK

The primary objective of social protection spending is to reduce current income poverty or income inequality.[4] This is achieved through transferring income from higher-income to lower-income ('poor') households.[5] Spending inefficiency therefore arises when the impact of spending on poverty is lower than what could be achieved via the most efficient allocation of a given spending envelope.[6] Below, we start by first developing the concept of spending efficiency within a simple poverty alleviation framework. We then show how this approach can be captured as a special case in a more general social welfare framework and discuss how this approach can be easily mapped to the widely used approach based on inequality impacts.

2.1 Imperfect Targeting

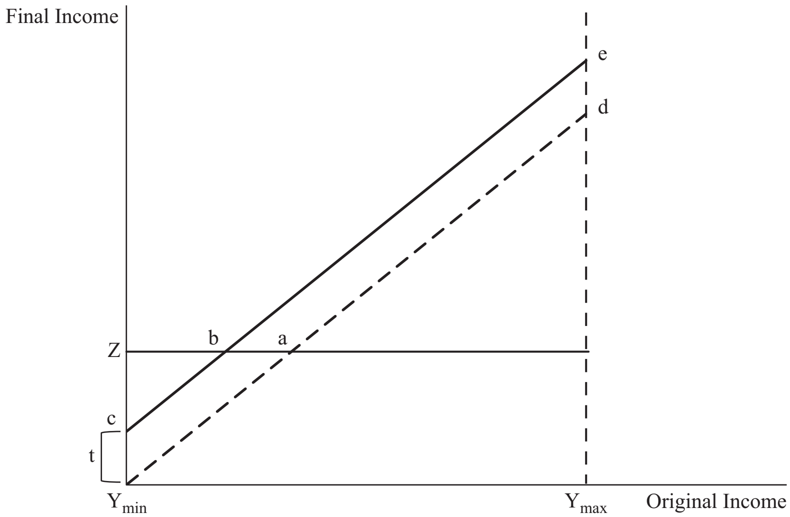
In general, for a given spending envelope (or transfer budget, B), the actual impact on poverty (P) can be written as:

$$dP \equiv \lambda.B \tag{14.1}$$

where λ can be interpreted as a measure of spending efficiency, that is, the poverty impact per unit of the budget reflecting how the budget is allocated across households. This measure of spending efficiency can be compared to different benchmarks, ranging from the maximum efficiency that could be achieved with a given budget[7] to observed efficiency levels in other countries with similar spending levels.

The source of targeting inefficiency can be seen from examining the *perfect* (first-best) transfer outcome while abstracting from household labor supply responses.[8] In the presence of a budget constraint, an ability to channel (or 'target') transfers to 'poor' (or lower-income) households will result in a greater decrease in poverty (and inequality). For instance, suppose we have a fixed transfer budget that is just sufficient to eliminate poverty, and perfect information on actual household income before any transfers (that is, 'original income,' for now assumed fixed). Income after transfers is 'final income.' Maximum and minimum household incomes are y_{max} and y_{min}, respectively, and z is the poverty line (Figure 14.1). The line dy_{min} shows that, before the transfer program is in place, households' final incomes are equal to their original incomes.

The *perfect* transfer scheme is one that gives a transfer to all poor households only (that is, those with income less than z), with transfer levels equal to their individual 'poverty gaps' given by the distance between their original income and the poverty line za.[9] This transfer program brings all poor households up to the poverty line and all non-poor households have equal final and original incomes. The poverty budget is thus represented by the area zay_{min} and this is also the minimum budget required to eliminate poverty. With reference to (14.1) above, for this transfer scheme, spending efficiency equals unity ($\lambda = 1$) since the poverty impact (dP) equals the total poverty gap (that is, poverty is eliminated) and the budget B also equals the poverty gap.[10]



Source: Based on Besley and Kanbur (1993).

Figure 14.1 Benefits of perfect means-tested targeting

Now, consider the case of a uniform (non-targeted) transfer program. This gives a uniform transfer t (= $c - y_{min}$) to all households, both poor and non-poor. Because of the 'leakage' of transfers to non-poor households, the transfers to poor households are no longer sufficient to eliminate all poverty. There are two forms of 'inefficiency' associated with this uniform transfer: (i) non-poor households receive a transfer, and (ii) some poor households (those in the line interval *ba*) receive transfers greater than their poverty gaps. As a result, with a fixed budget, the poorest households receive lower transfers and remain poor so that the poverty impact of the uniform transfer scheme is less than that of the perfect transfer scheme; less by the area *zcb*. The total 'leakage' of the budget (reflecting the two sources of inefficiency identified above) is given by the area *bade*, which for a fixed budget must also equal the area *zcb* representing the level of poverty after the uniform transfer program. Therefore, imperfect targeting (or the presence of 'targeting errors'[11]) results in a lower poverty impact for a given budget. In terms of (14.1) above, the reduction in the poverty gap (dP) will be less than the budget (B) so that spending efficiency is less than unity ($\lambda<1$).

2.2 Social Welfare

The above focus on poverty can be interpreted as a special case within a more general social welfare framework. Consider an economy with two groups; households and the government. Abstracting from behavioral responses,[12] let y_0 be household market income (that is, income before direct taxes and transfers) and y_1 be household income after direct income transfers) so that:

$$y_1 = y_0 + m \qquad (14.2)$$

where *m* denotes transfers. Let social welfare be described by a standard Bergson-Samuelson function of household welfare:

$$W(\ldots,V^h(\boldsymbol{p},y^h),\ldots) \tag{14.3}$$

where $V^h(.)$ is the indirect utility function of household *h* and **p** is a vector of commodity and factor prices facing the household (henceforth assumed fixed). The social welfare impact of a given transfer program with $dy^h = dm^h$ is:

$$dW = \sum\nolimits_h \frac{\partial W}{\partial V^h}\frac{\partial V^h}{\partial m^h}dm^h = \sum \beta^h \mathrm{d}m^h \tag{14.4}$$

where β^h is the social valuation of extra income to household *h*, the so-called social 'welfare weight.' Let the total transfer budget equal $B = \Sigma_h \mathrm{dm}^h$ so that (14.4) can be rewritten as:

$$\mathrm{dw} = \frac{\sum_h \beta^h \mathrm{dm}^h}{\sum_h \mathrm{dm}^h}\mathrm{B} = \mathrm{B}\sum\nolimits_h \beta^h\theta^h = \lambda.B \tag{14.5}$$

where θ^h is the share of the total budget received by household *h* and λ is the so-called *distributional characteristic* capturing the social welfare impact of a unit transfer delivered through the program (Diamond, 1975; Coady and Skoufias, 2004), which can be interpreted as an indicator of spending efficiency (or transfer progressivity). The greater the proportion of the budget ending up in the hands of lower-income households (that is, those with relatively high β^h), the higher the distributional characteristic and thus spending efficiency. Clearly λ can differ across transfer programs when welfare weights differ across households and the distribution of transfers differs across programs. Note also that the distributional characteristic is scale neutral in that it does not change in response to a scaling up or down of transfer levels.

The calculation of λ requires specifying social welfare weights. A very useful and common method for specifying these derives from Atkinson's (1970) constant elasticity social welfare function with the (relative) welfare weight of household *h* calculated as:

$$\beta^h = \left(\frac{y_k}{y_h}\right)^{\varepsilon} \tag{14.6}$$

where *k* is a reference income level (for example, mean income) and ε captures one's 'aversion to inequality' with this aversion increasing in ε.[13] For example, a value of $\varepsilon = 0$ implies no aversion to inequality (that is, a dollar is a dollar no matter to whom it accrues) so that all welfare weights take on the value unity. A value of $\varepsilon = 1$ implies that if household *h* has twice (half) the income of household *k* then its welfare weight is 0.5 (2.0) as opposed to unity for *k*. A value of $\varepsilon = 2$ similarly implies a welfare weight of 0.25 (4.0) for *h*. In practice, it is common to use $\varepsilon = 1$ as a reference (Chetty, 2006; Atkinson and Brandolini, 2010).

The poverty framework outlined earlier can be interpreted as a special case of this more general social welfare framework. For instance, if we set the welfare weights for poor households (for example, those in the bottom three income deciles) equal to unity and zero otherwise, then λ will become the share of transfers accruing to the poorest three deciles of the population, a measure of spending efficiency. If all transfers accrue to these poor households (that is, no leakage to the non-poor) then the perfect (or first-best) transfer program will attain the maximum level of spending efficiency with $\lambda = 1$.

To compare the efficiency of spending and the overall welfare impact of transfers across countries, we can write the total welfare impact of a transfer program in country *j* (dW_j) with budget B_j as:

$$dW_j = \lambda_j \cdot B_j \tag{14.7}$$

This can be usefully rewritten in percentage terms as:

$$\frac{dW_j}{Y_j} = \lambda_j \cdot \frac{B_j}{Y_j} = \lambda_j \cdot \tau_j \tag{14.8}$$

where Y_j is total national income and τ_j is the ratio of the transfer budget to total income. The percentage increase in welfare due to the transfer program in country *j* can then be compared to the increase for another country, and these differences can be decomposed into an efficiency component (λ_j) and an adequacy component (τ_j). In the illustrative empirical application below, we will discuss these components separately and their combined impact.[14]

It is useful to note that spending efficiency is simply a weighted average of social welfare weights β^h with θ^h, the share of transfers accruing to household *h*, as weights. Therefore, efficiency can be higher in one country compared to another because: (i) the share of transfers going to lower-income households with relatively high welfare weights is greater, or (ii) pre-transfer market income inequality is higher so that the relative welfare weight of lower-income households is higher and there is thus a greater social return to targeting lower-income households (Coady et al., 2021). If two countries have the exact same targeting outcomes (for example, the same share of transfers accruing to each income decile) then the country with higher market income inequality will show up as having higher efficiency. It is therefore important to clarify the respective roles of each of these factors in determining differences in spending efficiency across countries. One way of doing this is to compare the results for different formulations of welfare weights, one where welfare weights are continuous and dependent on the initial distribution of relative incomes (such as in the case of Atkinson's social welfare weights) and another where welfare weights do not depend on initial relative income levels, for example, the poverty-based welfare weights discussed above where the bottom three deciles have a welfare weight of one while higher-income deciles have welfare weights set at zero.[15]

2.3 Inequality

The literature on fiscal redistribution has traditionally been anchored in the literature on income inequality. Typically, the Gini coefficient for income after taxes and transfers (that is, disposable income) has been compared to that before taxes and transfers (that is, market income) to determine the extent of fiscal redistribution. This latter approach can be motivated by the social welfare framework used above as follows (Deaton, 1997).

Consistent with (14.3) above, let social welfare, W, be described by a function of individual incomes y_i as:

$$W = V(y_1, y_2, \ldots \ldots, y_N) \tag{14.9}$$

where N is the number of individuals in the population, V(.) is Paretian so that it is increasing in individual incomes and W satisfies the principle of transfers which requires that:

$$\frac{dW}{dy_i} > \frac{dW}{dy_j} \quad \text{for } y_i < y_j \tag{14.10}$$

with social welfare weights decreasing with individual incomes. To relate this social welfare framework to the income inequality framework it is useful to choose a social welfare function that has social welfare measures in the same units as individual welfare (that is, income) so that a proportional change in incomes for everyone leads to an equal proportional change in social welfare. This will be the case if V(.) is homogenous of degree one or can be thus transformed by a monotone increasing transformation. In such a case, social welfare can be written as:

$$W = \mu V\left(\frac{y_1}{\mu}, \frac{y_2}{\mu}, \ldots\ldots, \frac{y_N}{\mu}\right) \tag{14.11}$$

where μ is mean income in the population, and V(1,1,. . . . .,1) = 1 so that social welfare equals mean income when income is distributed equally with everyone having mean income. Since, by the principle of transfers, social welfare reaches a maximum equal to mean income, social welfare will be less than mean income when the income distribution is unequally distributed.

The above welfare function can then be rewritten as:

$$W = \mu\,(1 - I) \tag{14.12}$$

where (1 – I) is a scalar version of V(.) and I represents a measure of income inequality ranging from zero to unity (such as the Gini coefficient or Atkinson index).[16] In this case, I can be interpreted as the social welfare cost of inequality, that is, the loss in social welfare due to incomes being unequally distributed. Thus, any inequality index can be interpreted within a social welfare framework and, if it satisfies the principle of transfers, will be consistent with the welfare framework set out above. It should be noted that any such inequality index will have an implicit set of social welfare weights, for example, the Gini coefficient has an implicit set of welfare weights that depends on household rankings (Deaton and Muellbauer, 1983; Coady et al., 2021).

3. OTHER TARGETING COSTS

The above conceptual framework for understanding spending efficiency provides a starting point for informing real life choices made by policy makers wishing to address poverty and inequality through social protection spending. However, in practice, the appropriate level of efficiency will reflect not only social preferences (for example, as expressed in social welfare weights) but also the range of policy instruments available to policy makers. The latter in turn reflect the costs (real or perceived) they face in implementing these alternative policy instruments and how policy makers balance these costs against the benefits from achieving greater transfer progressivity (that is, spending efficiency). These costs include:

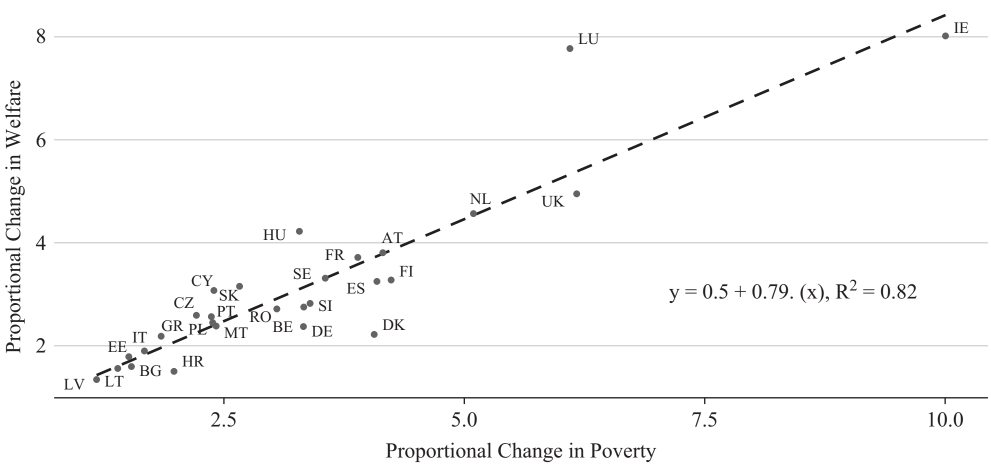
- *Moral hazard costs*: The means testing of benefits to avoid leakage of any transfers to non-poor households, which is inherent in the perfect targeting outcome, creates an extreme moral hazard since the withdrawal of benefits as other household income increases introduces a very strong work disincentive for both the poor and non-poor (essentially a 100 percent marginal tax rate). This can result in lower earnings income and higher transfers without a corresponding decrease in income poverty. Therefore, in terms of (14.1) above, since in the presence of moral hazard transfers increase without a one-for-one decrease in the poverty gap, spending efficiency will again be below unity ($\lambda<1$) when moral hazard is accounted for. The theory of *optimal* income taxation demonstrates that the optimal balancing of targeting and moral hazard costs requires less than full withdrawal of benefits (Mirrlees, 1971; Piketty and Saez, 2013).[17]
- *Administrative costs*: Many countries do not have the capacity to implement the sophisticated non-linear means-tested transfer schemes assumed in the optimal income tax literature, which requires collecting and verifying information on individual and household incomes. This can be especially burdensome in low-income settings, reflecting low administrative capacity, a large 'informal' sector constituting small-scale and self-employment activities, individuals with multiple and volatile sources of income (including in-kind income) and poor or non-existent bookkeeping.[18]
- *Private costs*: The costs of individuals acquiring sufficient capacity to comply with (or understand) benefit eligibility rules may be deemed undesirable or prohibitive. Households can also incur private costs involved in taking up transfers. For example, workfare programs involve households incurring an opportunity cost in terms of forgone income opportunities. Queuing involves similar, though usually much smaller, opportunity costs. Households may also face monetary costs for obtaining certifications required for the program (for example, a national identity card, and proof of residency or disability) and for transportation to and from program offices.
- *Political costs*: It is often argued that excluding the middle classes may remove broad-based support for transfer programs and make them unsustainable if voter support determines the budget and is in turn determined by whether the voter benefits directly from the program. Reflecting this constraint, policy makers may place a higher weight on transfers to middle-income households than otherwise. On the other hand, depending on the country and political context, efficient targeting to ensure that only those in need receive benefits may actually increase political support from those who support it based on its indirect benefits to them of reducing poverty (such as a feeling of social justice, increased political and social stability or lower taxes).
- *Social costs*: These costs may arise when the targeting of poor households involves publicly identifying households as poor, which may carry with it a social stigma. If the poorest households do not take up the transfer as a result, then this decreases the effectiveness of the program at getting transfers into the hands of the poorest. Such issues obviously take on additional importance when one appeals to concepts of poverty such as 'capabilities' (Sen, 1988). Targeting errors, that is, excluding some poor and including some non-poor, can also create social tensions within communities.

4. EMPIRICAL ANALYSIS

The above provides a transparent and practicable conceptual framework for undertaking an empirical analysis of the efficiency of social protection spending. Such an analysis should start by identifying the social outcomes of primary interest against which social protection spending can be evaluated. As indicated earlier, these include both measures of income poverty and income inequality. The analysis below is intended to illustrate how such an evaluation can be developed for advanced countries using data for European countries from Euromod, which are based on microsimulation studies informed by household survey and administrative data. We also discuss how the approach set out in this paper corresponds to the approach commonly used to evaluate public spending inefficiencies more generally, namely, the use of stochastic frontier regressions. We argue that, in the context of social protection spending, the latter do not appear to provide a useful empirical approach to evaluating spending efficiency; essentially, they only appear to work well when one has access to data typically generated through microsimulation studies, in which case they are effectively redundant.

The analysis is based on databases available on the Euromod website. These databases provide information on direct taxes and transfers for 28 EU countries broken down by income deciles (see Appendix Table 14.1 for an example of the data available). Together with data on average decile per capita incomes, this information is sufficient to calculate the parameters of interest. We focus on data for 2016, which is the most recent year available. Benefits include social insurance (for example, pensions) and social assistance cash transfers (means tested and not means tested), while taxes include social contributions (or payroll taxes) and personal income taxes. To illustrate the application of the above conceptual framework for evaluating spending efficiency, we focus only on social assistance transfers.

Figure 14.2 presents the poverty and social welfare impact of transfers based on (14.8) above. The results are based on two sets of social welfare weights. First, welfare weights are set equal to one for the poorest three income deciles (the 'poor') so that λ becomes the share of



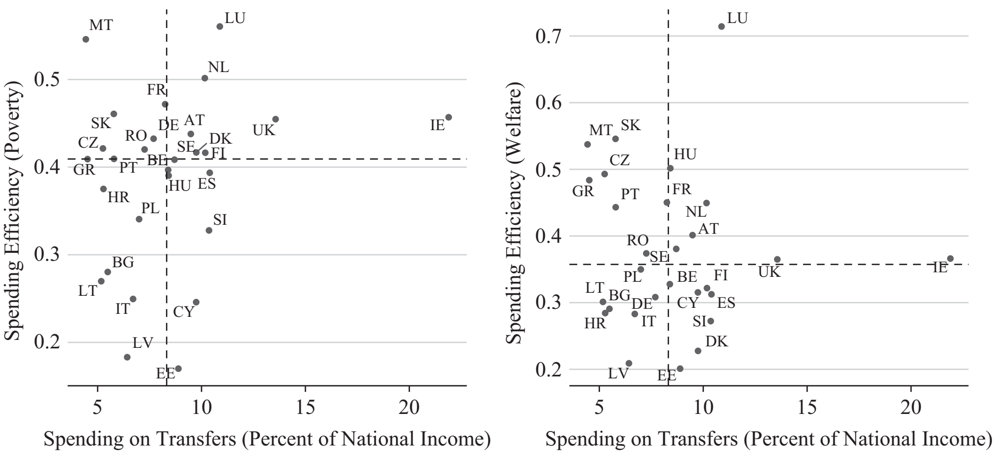
Source: Authors' calculations based on Euromod data.

Figure 14.2 Poverty and welfare impacts of transfers in European countries, 2016 transfers accruing to the poor and the social welfare impact becomes the share of total national income accruing to the poor. We refer to this simulation as the 'poverty' impact. Second, welfare weights are calculated as the ratio of each decile's mean income to the mean income of the bottom decile. Therefore, λ will increase with the share of transfers accruing to the bottom deciles and will reach a maximum of $\lambda = 1$ when all transfers accrue to the poorest decile. We refer to this simulation as the 'welfare' impact.

The results show substantial variation in the poverty and welfare impacts across countries. However, the correlation between both impacts is very high, with a total explained variation of 82 percent and most of the ranking changes being between countries with similar levels of impact.[19] Among the biggest re-rankings are Croatia, Germany and Denmark, which do relatively well in terms of poverty impact but less well in terms of welfare impact. Since both poverty and welfare impacts reflect the product of spending efficiency and spending adequacy, and adequacy is fixed for a given country, these re-rankings reflect the difference between efficiency measures across the poverty and welfare impacts.

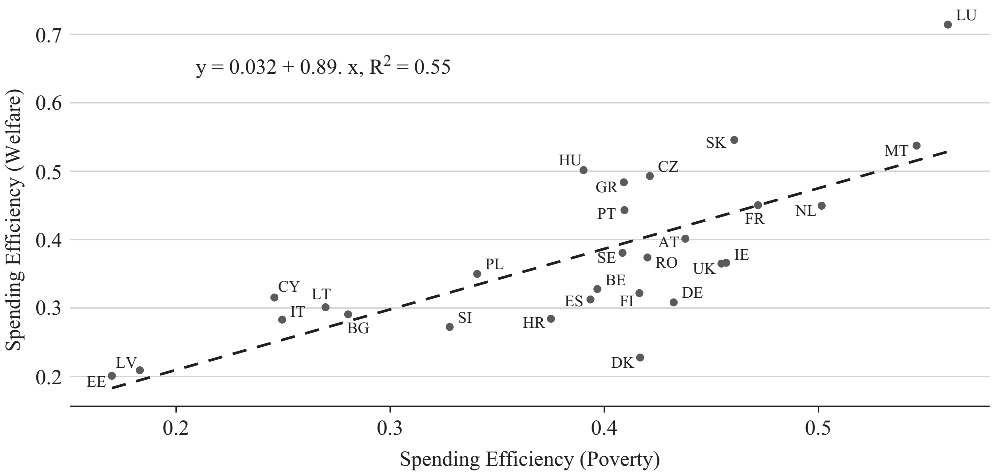
Figure 14.3 plots spending efficiency and adequacy against each other for both impacts. In general, the correlation between spending adequacy and spending efficiency is quite weak across both impacts. Some countries combine high adequacy with high efficiency, resulting in high overall poverty and welfare impacts (Luxembourg, Ireland and Netherlands). Other countries combine low adequacy with low efficiency resulting in low overall impacts (Latvia, Lithuania, Bulgaria and Italy). In other countries, low adequacy is offset by high efficiency (Malta, Slovakia, Czech Republic and Greece), while others have the opposite (Slovenia, Cyprus and Estonia).

For both poverty and welfare impacts there is substantial variation in spending efficiency across countries (Figure 14.4). Overall, there is a strong correlation between poverty and welfare spending efficiency with countries having relatively high (low) poverty efficiency also tending to have relatively high (low) welfare efficiency (Figure 14.4). High efficiency countries under both measures include Luxembourg, Malta, Netherlands, France and Slovakia,



Source: Authors' calculations based on Euromod data.

Figure 14.3 Spending efficiency and adequacy for European countries, 2016



Source: Authors' calculations based on Euromod data.

Figure 14.4 Poverty and welfare spending efficiency

while low efficiency countries include Estonia, Latvia, Bulgaria and Italy. The cases of Denmark and Hungary are instructive. While both are in the middle of the distribution for poverty efficiency, under welfare efficiency Denmark moves to the bottom of the distribution while Hungary moves towards the top. This reflects the relatively high 'welfare cost of leakage' in Denmark because of high market income inequality compared to the low costs in Hungary due to low market income inequality.

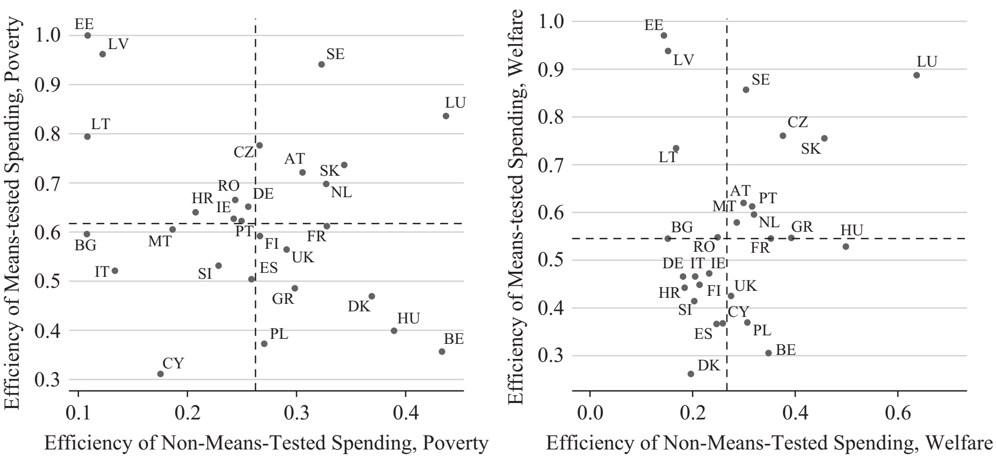
The above framework and analysis of spending efficiency provides a basis for discussing the potential of regression-based approaches for evaluating spending efficiency in the context of social protection spending. For example, it is common in analyses of public spending efficiency more generally to use *stochastic frontier* (SF) analysis (see Annex I for a more detailed discussion) to estimate spending efficiency. This typically involves regressing some measure of the 'outcome,' which the spending was intended to enhance, for each country on spending levels of these countries. Other determinants of the outcome are also included as explanatory variables to avoid attributing their effects to the impact of spending on the outcome of interest or to variations in spending efficiency.

In the context of social protection spending, social welfare in (14.12) above can be taken as the appropriate outcome variable of interest. But, since we are focusing on spending efficiency and abstracting from differences in mean incomes across countries, we can simply focus on (1 – I) as our measure of social welfare. The corresponding regression analysis would have (1 – I) as the dependent variable, with I set at post-transfer inequality and transfer spending as an independent variable. To control for factors other than transfer spending that affect post-transfer I we would need to effectively control also for I before transfer spending (that is, pre-transfer I). However, when we apply a standard SF analysis to these data we find large discrepancies between our measures of spending efficiency, which are derived from a well-defined social welfare framework, and the estimates of efficiency produced by the SF analysis. We therefore conclude that, in the context of social protection spending efficiency, SF analysis can produce misleading results, possibly reflecting that its underlying assumptions on the distribution of efficiency are too rigid. In any case, since one needs to control for pre-transfer I, one already has sufficient information to calculate the efficiency estimates directly as the difference in pre-transfer and post-transfer I divided by total transfer spending as per (14.1) above, thus making SF analysis redundant.

Regression analysis, however, does provide a useful empirical framework for identifying the relative importance of spending efficiency and spending adequacy in determining variations in the welfare impact of fiscal redistribution using the results discussed above. Based on (14.7) above, a linear regression of the welfare impact (dW) on efficiency (λ) and adequacy (τ) allows for an approximation of the independent impacts of each component using ANOVA analysis.[20] Focusing on the welfare impact discussed above, as expected the explained variation of such a regression is very high at 98 percent. Of this, variations in efficiency across countries account for one third of the variation on overall welfare impact and variation in adequacy accounts for the remaining two-thirds. The corresponding figures for our poverty impact are 24 and 75 percent.

5. POLICY IMPLICATIONS

Estimates of spending efficiency can be used to identify the potential each country has for enhancing the impact such spending has in achieving underlying social objectives. In general, a country with relatively low spending efficiency can enhance efficiency in various ways. It can increase the share of spending allocated to better targeted transfer programs, such as means-tested programs. Second, it can improve the targeting of these means-tested programs where this is relatively low. Third, it can improve the targeting of non-means-tested transfer programs where this is low. We will illustrate each of these below using the empirical analysis above but, for presentational convenience, will focus only on welfare efficiency. Figure 14.5



Source: Authors' calculations based on Euromod data.

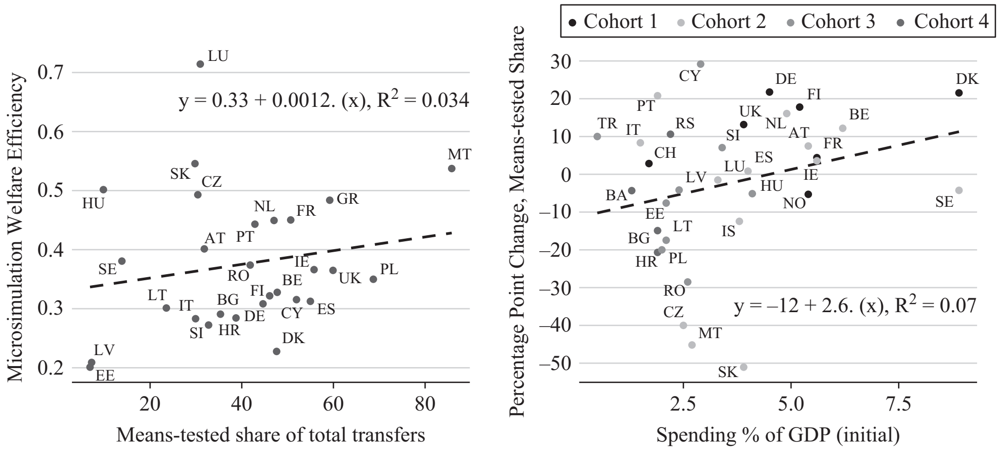
Figure 14.5 Spending efficiency for means-tested and non-means-tested transfers shows that the efficiency of spending is typically much higher for means-tested programs with a median above 0.5 compared to a median of below 0.3 for non-means-tested programs. But it is also the case that there is substantial variation in efficiency across countries for both sets of programs.

The discussion below is meant to be instructive as to how to move from an efficiency analysis to a discussion of policy considerations. Actual policy prescriptions for any country would need to be informed by more country-specific and program-specific analyses and broader policy considerations. For instance, as indicated earlier, countries must balance the benefits of improved targeting against various costs associated with better targeting—administrative constraints may preclude the use of sophisticated mean-tested transfer programs; political considerations may require including middle-income groups in transfer programs to generate support for program budget allocations; or means testing to narrow 'poor' household groups may generate social stigma and possibly low take-up of benefits. In addition, countries need to be able to contain the moral hazard costs arising from the work disincentives inherent in means-tested transfers. The fact that even advanced European economies with significant administrative capacity have a high share of social assistance transfers that are not means tested suggests that many of these forces are at play in practice.

5.1 Enhancing Role of Means-Tested Benefits

All European countries have some sort of means-tested Guaranteed Minimum Income (GMI) program (Immervoll, 2009; Frazer and Marlier, 2016; Crepaldi et al., 2017; Coady et al., 2021). However, the share of total assistance transfers that is means tested varies substantially across countries. One option then for countries to consider is whether they want to integrate some of existing non-means-tested social assistance spending into an existing means-tested system. The share of means-tested benefits in countries with relatively high initial spending levels has trended upwards, increasing by an average of 25 percent between 1990 and 2017 (Figure 14.5), presumably partly to contain high spending levels. Other countries with relatively low initial spending levels have increased the share of non-means-tested transfers as they have increased spending levels, presumably partly to expand coverage of benefits in the context of limited administrative capacity.

The share of means-tested benefits is particularly low in Estonia, Latvia, Hungary, Serbia and Lithuania. Among these, Estonia, Latvia and Lithuania have especially low spending efficiency and could therefore consider expanding the role of means-tested programs. Some countries with a relatively high share of means-tested benefits still appear to have relatively low spending efficiency. These include Poland, United Kingdom and Spain. These countries can consider how to strengthen targeting of these programs, for example, through adopting a more rapid withdrawal of benefits. However, this should be accompanied by measures that help to contain work disincentives (Annex II). In addition, measures should be taken to ensure that the administration of such benefits is efficient and that the private costs of potential beneficiaries acquiring knowledge of the program and applying are not high. Stigma concerns can possibly be addressed through linking benefit receipt to the receipt of formal or on-the-job training. Emphasizing the administrative efficiency, low fiscal cost and work-focused dimensions of these programs may also help generate political support.



Note: Cohorts are defined according to the earliest date for which data are available: (1) Denmark, Finland, France, Norway, Switzerland, Germany and United Kingdom; (2) Italy, Ireland, Netherlands, Luxembourg, Austria, Sweden, Portugal, Belgium, Spain, Czech Republic, Malta, Slovakia and Iceland; (3) Lithuania, Slovenia, Latvia, Estonia, Hungary, Cyprus, Romania and Turkey; and (4) Bulgaria, Croatia, Serbia and Bosnia and Herzegovina.

Sources: ESSPROS Dataset; IMF Staff Calculations.

Figure 14.6 Spending efficiency and share of means-tested transfers

5.2 Increasing the Efficiency of Non-Means-Tested Benefits

An alternative to increasing the share of means-tested (MT) benefits is to improve the targeting of non-means-tested (NMT) benefits. In general, NMT benefits are less targeted than MT benefits (Figure 14.6) and countries with a high share of NMT benefits that are badly targeted (such as Estonia, Latvia and Lithuania) can consider improving their targeting. This can be done both through changing eligibility rules and benefit levels. Universal child benefits provide a useful example. These benefits are available to all children below a certain age. Since, in general, not all children live in lower-income households, there will typically be substantial leakage of benefits to higher-income households with children. Enhancing the targeting of child benefits can be done through a combination of changing eligibility rules and varying benefit levels when the leakage of benefits to higher-income groups (that is, the distribution of children across the income distribution) varies by age group. Benefit eligibility could be restricted to age groups where leakage is relatively low, or benefit levels could be higher for these groups. Or benefits for age groups with greater leakage could be means tested, including by reducing benefit levels as household income increases. Alternatively, ensuring that benefit income is included in the tax base within a progressive income tax system will enhance the progressivity of the overall transfer and tax system, even if benefit efficiency is not affected.

6. CONCLUDING REMARKS

As with other types of public spending, an evaluation of the efficiency of social protection spending needs to be firmly anchored in a clear set of social objectives that it is intended to achieve. In the context of social protection spending ('transfers'), these are typically specified as decreasing income poverty and income inequality. This chapter sets out a well-established conceptual framework to guide the evaluation of social protection spending efficiency consistent with these social objectives.

Within this framework, spending efficiency is seen to be synonymous with the more frequently used term of 'spending progressivity,' which ceteris paribus increases the more spending is concentrated (targeted) on the lowest income ('poorest') groups. However, efficiency will also depend on the inequality of pre-transfer market income. For any distribution of transfers, efficiency will be higher with greater market income inequality, reflecting higher social returns to targeting. Therefore, if two countries have the same distribution of benefits across households (that is, the same targeting) then the country with higher initial inequality will have higher spending efficiency. A comprehensive empirical analysis of spending efficiency should therefore identify the respective roles of targeting outcomes and the social returns to targeting in determining overall spending efficiency. This can be done through considering alternative specifications of welfare weights. Where initial inequality (and the social returns to targeting) are relatively high, consideration should also be given to reducing disposable income inequality through measures that reduce market income inequality.

In the presence of costs associated with targeting, maximum efficiency is neither optimal nor indeed attainable. These costs include moral hazard, administrative, social, political and private costs. The optimal level of efficiency will therefore differ across countries when the costs of targeting differ and when there are differences in countries' preferences over how best to balance the benefits and costs of targeting. Care therefore is needed when using the results of spending efficiency analysis to inform policy choices, which also requires a more detailed analysis of the different components of total transfers and the (real or perceived) costs associated with targeting in a specific country context.

Reflecting the above, it is important that spending efficiency analysis incorporates the following features to strengthen its ability to inform policy choices:

- An evaluation of efficiency across a common set of social preferences over income inequality to see how sensitive relative spending efficiency is to different preferences. The social welfare framework has the advantage that these preferences are more easily made explicit for the purposes of sensitivity analysis.
- A clear understanding of the role played by the different components of efficiency (that is, targeting and initial inequality) in explaining differences in overall efficiency. This can be achieved through undertaking sensitivity analysis over alternative social welfare weights.
- A clear understanding of the role played by the different components of transfers (for example, means-tested and non-means-tested transfers) in determining overall efficiency of transfers.

The insights from the above analyses would need to be combined with more detailed information on the design of existing programs to adequately inform policy choices.

Benchmarking overall spending efficiency across countries, and the contributions of different transfer components, can provide a very useful starting point for identifying the potential for strengthening efficiency in different countries. For instance, countries with relatively low efficiency can consider increasing the role of means-tested transfers, which are typically better targeted than other transfer components. In some countries there may be room for enhancing the targeting of benefits through a sharper withdrawal of benefits as household income increases. Or countries can consider enhancing the targeting of non-means-tested transfers through more progressive eligibility rules and benefit levels. However, recommendations on how to enhance efficiency should also identify the trade-offs that are involved in terms of the various costs of targeting. For example, increased reliance on means-tested transfers may involve higher administrative and moral hazard costs, or increased reliance on better-targeted non-means-tested transfers may involve different social and political costs. In the end, it is country policy makers who should decide the appropriate level of targeting efficiency reflecting social preferences and the various constraints they face when choosing and implementing alternative policy instruments.

Appendix Table 14.1 EUROMOD: Ireland distribution of income, taxes and transfers

Decile Group	Disposable Income	Original Income	Means-Tested Benefits	Non-Means-Tested Benefits	Public Pensions	All Taxes	Social Insurance Contrib. (SICs) (2)	Simulated Benefits, of All Benefits (%)	Simulated Taxes, of All Taxes (%)
1	1,099.8	255.6	657.8	236.2	44.4	64.1	30.2	92.2	100.0
2	1,586.3	354.9	951.8	265.1	84.5	58.1	12.0	92.5	100.0
3	1,930.4	689.5	922.9	275.1	155.9	70.7	42.4	90.4	100.0
4	2,289.1	1,268.5	706.5	355.1	177.7	136.5	82.2	87.6	100.0
5	2,738.9	2,014.4	413.0	341.7	360.0	253.1	137.1	88.3	100.0
6	3,208.0	2,960.4	173.5	376.7	356.5	462.8	196.2	89.7	100.0
7	3,697.3	3,708.7	112.1	428.4	376.0	667.9	260.0	89.1	100.0
8	4,125.0	4,652.5	64.3	333.2	405.0	998.4	331.7	91.5	100.0
9	5,076.7	6,445.4	22.3	348.5	268.3	1,525.7	482.1	90.5	100.0
10	6,993.0	10,726.6	14.7	243.1	148.0	3,345.3	794.0	88.1	100.0
All	3,264.8	3,304.1	409.1	318.1	233.7	763.4	236.8	90.1	100.0
Poor (3)	1,283.2	270.2	790.2	249.9	53.6	58.5	22.1	92.7	100.0

Definitions

original income	employment income + investment income + income of children under 16+ private pension + income from property + private transfers received + self-employment income + pension from other employment + pension from public sector
taxes (sim.)	personal income tax + universal social charge + household charge - mortgage interest relief
taxes (data)	property tax
employee SICs (sim.)	employee PRSI + superannuation + public sector pension related deduction
self-empl. SICs (sim.)	self-employed PRSI + self-employed investment and rental income SIC
benefits (sim.)	maternity benefit + state pension (non-contributory) + one parent family payment + widows non-contributory pension + disability allowance + illness benefit + supplementary welfare allowance + family income supplement + jobseekers benefit + jobseekers allowance + injury benefit + child benefit + state pension (contributory) + state pension
benefits (data)	rent and mortgage supplements + fuel allowance + minor social assistance benefits + residual family allowances + grants/education (training) allowances + education grant (from FÁS) + household benefit package + non-Irish social

Notes:
1. The categories of income components chosen for these tables are simply for illustrative purposes. The categorization of instruments is an area where EUROMOD offers a high degree of flexibility which is needed if results are to conform to different conventions and are to be used for a range of purposes. June 2011-2016 market exchange rates are used for non-euro countries.
2. Social insurance contributions refer here to the sum of employee and self-employed contributions and all benefits also include public pensions.
3. Poor: households at risk of being in poverty, i.e., with equivalized disposable income below 60% of the median.

Source: EUROMOD data available at: https://www.euromod.ac.uk/using-euromod/statistics

ANNEX I. COMPARISON OF MICROSIMULATION AND STOCHASTIC FRONTIER ANALYSIS

A common empirical framework for assessing efficiency is stochastic frontier analysis (SFA). Originally developed in the context of estimating the efficiency of firm-level production, SFA can be applied to a wide range of settings where efficiency is an area of interest, including public spending. In this annex we discuss how such a SFA could be applied to an analysis of social protection spending efficiency and benchmark its efficiency estimates to the results presented in the paper. This comparison suggests that the SFA approach may result in misleading efficiency estimates, which presumably reflects the relatively rigid assumptions that are required as regards the distribution of (unobserved) efficiency and random error. In any case, we argue that once one has adequate information to control for other factors that affect outcomes, this is likely to be adequate to undertake the type of microsimulation analysis presented in the paper.

The benchmark SFA model can be expressed as:

$$\boldsymbol{y_i} = \boldsymbol{F}(\boldsymbol{x_i}, \boldsymbol{\beta})\boldsymbol{E_i} \tag{A14.1}$$

where $\boldsymbol{i}$ indexes the observation unit (for example, firm or country), $\boldsymbol{y_i}$ is $\boldsymbol{i}$'s output (or change in outcome), $\boldsymbol{F}(\cdot)$ is a production function with inputs $\boldsymbol{x_i}$ and $\boldsymbol{E_i}$ is an unobserved efficiency term with $0 \leq \boldsymbol{E_i} \leq 1$. Taking the natural logarithm, results in the following linear specification:

$$\ln(\boldsymbol{y_i}) = \boldsymbol{\alpha} + \boldsymbol{\beta}\ln(\boldsymbol{x_i}) + \ln(\boldsymbol{E_i}) \tag{A14.2}$$

Letting $\ln(\boldsymbol{E_i}) = \boldsymbol{v_i} - \boldsymbol{u_i}$, the efficiency term can be decomposed into a standard random shock parameter $\boldsymbol{v_i} \sim \boldsymbol{N}(0, \boldsymbol{\sigma}^2)$ and an $\boldsymbol{i}$ -specific technical inefficiency parameter, $\boldsymbol{u_i} \geq 0$. This composite error structure allows for the existence of shocks that are unrelated to technical inefficiency and therefore controls for any distortionary effect they may have on assessments of efficiency. Model estimation requires assumptions of the distribution of $\boldsymbol{u_i}$ for which a common baseline assumption is the half-normal distribution $\boldsymbol{u_i} \sim \boldsymbol{N}^+(0, \sigma^2_{ui})$ which we also use here. This model is then estimated using maximum likelihood methods.

In the context of social protection spending, $\boldsymbol{y_i}$ corresponds to the welfare impact of spending consistent with equation (14.1) in the paper:

$$\mathrm{dW} \equiv \lambda.\mathrm{B} \tag{A14.3}$$

where B is social protection spending for each country, λ is spending efficiency and W is social welfare which can be evaluated as:

$$\mathrm{W} = \mu\,(1 - \mathrm{I}) \tag{A14.4}$$

where I represents a measure of income inequality and μ is mean country income. Substituting (A14.4) into (A14.3) we get:

$$\mathrm{dW} = \mathrm{dW}_1 - \mathrm{dW}_0 = (1 - \mathrm{I}_1) - (1 - \mathrm{I}_0) \tag{A14.5}$$

where sub-scripts denote the values taken by variables before (0) and after (1) transfer spending, and we drop the term for mean country income since we are focusing on spending efficiency and abstracting from changes in country mean income.

The standard approach in SFA applied to spending efficiency is to use the outcome variable after spending W_1 as the dependent variable and spending B as the explanatory variable. This specification requires that the regression also includes additional explanatory variables that affect the outcome variable independently of B. The corresponding regression for analysis of social protection spending efficiency is then:

$$(1 - I_1) = \alpha + \beta\, B + \gamma\, (1 - I_0) \qquad \text{(A14.5)}$$

In the analysis presented below, consistent with the social welfare approach used in the paper, we use the Atkinson measure of inequality for two values of the inequality aversion parameter (that is, $\varepsilon = 1$ and $\varepsilon = 2$).

Appendix Table 14.2 presents the regression results for different model specifications, the first panel for $\varepsilon = 1$ and the second for $\varepsilon = 2$). In Model 1 the dependent variable corresponds to the welfare *impact* of transfers as calculated by equation (14.8) in the paper and the single dependent variable is the level of the transfer budget. This specification is therefore the closest to the analysis presented in the paper embedded with an SFA. The estimated coefficient indicates that, on average, a 10 percent increase in the transfer budget is associated with an 8.6–9.1 percent increase in the welfare *impact*, consistent with there being a low correlation between spending and efficiency in the data.

In Model 2 the dependent variable is as in equation (A14.5) above with only B included as an explanatory variable and the outcome variable is $(1 - I)$. The coefficient on B indicates that a 10 percent higher level of spending is associated with 0.7–1.9 percent increase in post-transfer welfare. However, this measure fails to allow for the fact that initial pre-transfer inequality also varies across countries in a manner correlated with country budgets (for example, spending is higher in countries with initially higher pre-transfer inequality), so that the estimated coefficient is partly picking up this correlation. Model 3 corrects for this by including initial pre-transfer inequality as an additional explanatory. As expected, there is a strong positive relationship between pre-transfer and post-transfer inequality, with a narrower range of 1.1–1.4 on the budget variable, the latter higher estimate being consistent with a higher social welfare return to redistributive transfers the greater the aversion to inequality.

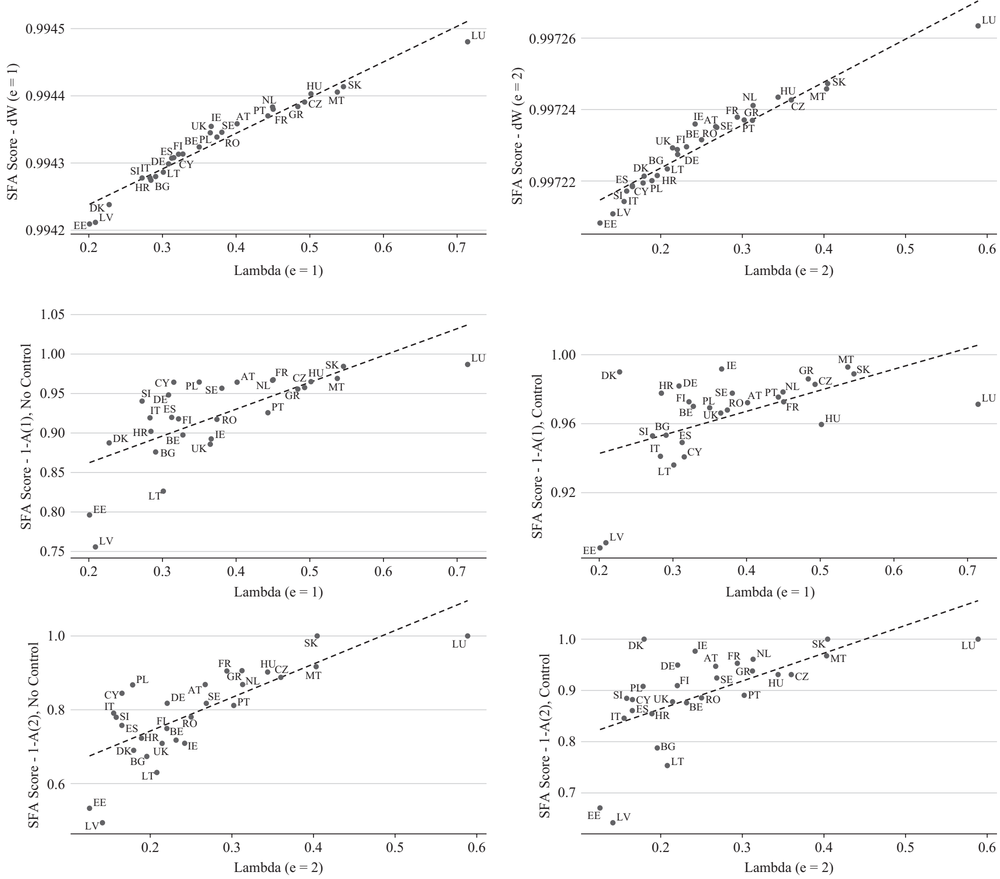
Finally, we compare the efficiency scores from the various SFAs to those estimated in the paper (Appendix Figure 14.1). The first panel shows this for Model 1 and not surprisingly, given the very close correspondence between both analyses, the efficiency estimates are extremely highly correlated—correlation coefficient of 0.96–0.97 and rank correlation of 0.99–1.00. However, the SFA estimates very high levels of efficiency across all countries compared to those estimated in the paper, which also vary a lot more across countries. This suggests that the SFA analysis is attributing much of the variation in efficiency to the random error term. The second panel presents efficiency estimates for Model 2 with the more traditional SFA specification but with only B as an explanatory variable. The correlation between efficiency estimates falls substantially—correlation coefficients of 0.70–0.75 and rank correlations of 0.75–0.78—and the ranking of countries by efficiency is substantially different than that presented in the paper. The third panel presents efficiency estimates for Model 3 when we add

Appendix Table 14.2 Results from the SFA regressions

	Panel A: Inequality Aversion, ε = 1		
	log (dW)	log (1-Post-Transfer A(1))	log (1-Post-Transfer A(1))
log (Transfer Budget. Percent GDP)	0.91***	0.07**	0.11***
	(0.16)	*(0.04)*	*(0.02)*
log (1-Pre-Transfer A(1))			0.58***
			(0.10)
log (Variance of Random Error)	–2.44***	–7.17***	–8.51***
	(0.27)	*(0.82)*	*(0.96)*
log (Variance of Technical Efficiency Term)	–9.89	–4.46***	–6.09***
	(271.24)	*(0.38)*	*(0.51)*
	Panel B: Inequality Aversion, ε = 2		
	log (dW)	log (1-Post-Transfer A(2))	log (1-Post-Transfer A(2))
log (Transfer Budget, Percent GDP)	0.86***	0.19***	0.14***
	(0.19)	*(0.00)*	*(0.00)*
log (1-Pre-Transfer A(2))			0.32***
			(0.00)
log (Variance of Random Error)	–2.08***	–38.09	–37.63
	(0.27)	*(611.23)*	*(0.13)*
log (Variance of Technical Efficiency Term)	–11.32	–2.43***	–3.67***
	(382.36)	*(0.27)*	*(0.35)*
Coverage	2016	2016	2016
Countries/Observations	28	28	28

Note: An analysis of the OLS residuals for Model's 1 and 2 find the residuals to be normally distributed with a leftward skew, suggesting the presence of technical inefficiency. We specify a heteroskedasticity robust estimation procedure for all frontier analysis. The analysis is based on a 28-country sample for 2016. Stars denote: $p < 0.01$ ***, $p < 0.05$ **, $p < 0.10$ *.

initial pre-transfer inequality as an additional explanatory variable. Somewhat surprisingly, the correlation falls even further—correlation coefficients of 0.54–0.61 and rank correlation of 0.53–0.72—and again country efficiency rankings differ substantially from the estimates in the paper. This suggests that the underlying assumptions in the SFA on the distribution of efficiency may be very inappropriate and are biasing efficiency estimates significantly.



Appendix Figure 14.1 Relationship between different SFA efficiency measures and lambda

ANNEX II. POLICIES FOR ADDRESSING LABOR MARKET DISINCENTIVES IN MEANS-TESTED PROGRAMS

A key challenge facing countries when designing means-tested benefit systems is how to contain the work disincentives associated with the withdrawal of benefits as household members enter employment and increase earnings. While a sharp withdrawal of benefits helps to better target benefits to lower-income groups (that is, enhance spending efficiency) and to contain fiscal cost, it also significantly reduces work incentives. Many countries address this trade-off by reducing benefit generosity while more gradually withdrawing benefits (OECD, 1997; Immervoll, 2009; Pena-Cassas et al., 2013; Coady et al., 2021). However, this comes at a high cost in terms of undermining the primary objective of poverty alleviation.

Countries can use various approaches to reduce work disincentives while combining high generosity with high benefit withdrawal rates. These include:

- *Income disregards*. This allows beneficiaries to earn a certain amount without any reduction in benefits, that is, a zero-withdrawal rate over this income range, after which higher withdrawal rates are applied. This strengthens work incentives, especially for low-income individuals who are often more responsive to financial incentives.
- *Fixed duration benefits*. This allows beneficiaries to earn income without withdrawal of benefits for a certain period after which high withdrawal rates are applied. This strengthens work incentives in the short term and allows time for the individual to develop work habits and skills and thus strengthen attachment to the labor market.
- *Benefit conditionality*. Recent reforms to means-tested benefits in many countries have focused on strengthening conditionality for receipt of these benefits to counteract the disincentive effects of sharp benefit withdrawal. These conditions involve regular reporting on work search, participation in training, and the requirement that recipients accept jobs that they have been offered or have been sourced for them, including possible government-sponsored jobs. Failure to engage with these programs is typically met with benefit sanctions

Careful monitoring of labor market developments of benefit recipients can also be used as signals of the need for further tailored public support or benefit conditionality such as training or specific job placements.

NOTES

1. We acknowledge the excellent research assistance provided by Riki Matsumoto.
2. This chapter does not address public pension spending, which is taken up in Chapter 8.
3. The Euromod project has produced extensive work combining EU-SILC household survey data with a standardized microsimulation model adapted to member countries to evaluate the distributional impact of tax and benefit systems and their reform (Sutherland and Figari, 2013; Avram et al., 2014). Caminada et al. (2017, 2019) have used Luxembourg Income Survey data for similar purposes, while the OECD also regularly produces estimates of fiscal redistribution based on household survey data (Immervoll and Richardson, 2011; OECD, 2008, 2011; Causa and Hermansen, 2019). Although analyses of fiscal redistribution and supporting databases for low-income and middle-income countries are much more limited, Inchauste and Lustig (2017) provide an overview of the growing literature in this area under the auspices of the Commitment to Equity (CEQ) project.
4. Other objectives could include 'promotion' (rather than 'protection') which involves enhancing the asset base of 'poor' households aimed at increasing future incomes or promoting attachment to the labor market. In practice, there may be a trade-off between these two social objectives so that the impact of spending on current poverty for a given spending budget may be lower than otherwise would be the case.
5. For convenience, we will often use the term poverty to denote both poverty and inequality.
6. We abstract from the issue of the determination of the optimal spending envelope (or transfer budget) since the empirical literature focuses primarily on efficiency given a specific level of spending. The empirical literature is also very diverse in terms of the welfare indicator used and the population group analyzed (for example, households, individuals and workers).
7. This could be calculated using a household survey containing information on household income and transfers received.
8. For more detailed discussion, see Atkinson (1995ab); Besley and Kanbur (1993); Coady et al. (2004a; 2004b); and van de Walle (1998).
9. Note that *perfect targeting* involves: (i) *Perfect beneficiary targeting* where only the poor receive transfers and all the poor receive transfers; and (ii) *Perfect transfer targeting* where all of the transfer budget goes to the poor and the poor receive exactly the gap between their incomes and the poverty line. We use the term 'perfect' to distinguish this scheme from the conventional notion of an 'optimal' scheme that minimizes the efficiency cost associated with means-tested transfers.
10. Note that this holds even when the transfer budget is less than the poverty gap (and therefore insufficient to eliminate poverty) as long as transfers are targeted at the poorest households first. In this case, the incomes of the poorest households would be brought up to a common level below the poverty line.
11. Targeting errors are often categorized as 'errors of exclusion' (that is, some poor do not receive transfers) and 'errors of inclusion' (that is, some non-poor receive transfers).
12. Most empirical papers on fiscal redistribution abstract from the important issue of behavioral responses arising from taxes and transfers. However, such responses could potentially be very important in deciding on the optimal level of fiscal redistribution since they generate an efficiency–equity trade-off (Piketty and Saez, 2013; Bargain, 2017). The presence of such responses also means that the level and distribution of 'original' incomes (that is, incomes prior to the imposition of taxes and transfers) may be different from the level and distribution of 'market' incomes (that is, 'disposable' incomes after taxes and transfers, minus taxes and transfers), the extent of these differences depending on the elasticity of income to net transfers and how this varies across income groups. While the conceptual framework used here applies regardless of whether original or market incomes are used in equation (14.2), the empirical results and their policy implications could, of course, be sensitive.
13. This approach is well established in the literature; for examples, see Atkinson and Stiglitz (1980); Newbery and Stern (1987); and Ahmad and Stern (1991). For a discussion of alternative approaches to setting welfare weights, see Kanbur and Tuomala (2011).
14. In Annex I, we also compare the efficiency measure to the types of efficiency measures that would emerge from the regression approaches often used in practice when discussing spending efficiency.
15. In the context of the Gini coefficient, which will be discussed briefly below and is commonly used in analyzing spending progressivity, this can be achieved through comparing results for the absolute change in the Gini due to transfers to the proportional change since, in relative terms, *certeris paribus* the latter will be smaller than the former for countries with high initial Gini inequality.
16. Note that, in the context of the iso-elastic social welfare function used by Atkinson (1970), this is equivalent to setting social welfare equal to the level of income which, if equally distributed, will give the same level of social welfare as the existing distribution of income, which he refers to as 'equally distributed equivalent' (EDE) income.
17. Simulations within a social welfare framework suggest marginal income tax rates (MTRs) of (60, 70, 90) percent on low incomes for inequality aversion parameters of (0.5, 1.0, 2.0), and corresponding optimal grants equal to (36, 54, 67) percent of median income (Coady and Le, 2020). Similar estimates were found for the UK by Brewer et al. (2010). Simulations within a poverty alleviation framework suggest optimal MTRs on low incomes of

around 60−70 percent (Kanbur et al., 1994). However, the literature also finds that very low benefit withdrawal rates for the lowest earners (or even wage subsidies, that is, negative benefit withdrawal rates) may be optimal when one takes account of disincentives to participate in the labor market in the first instance (that is, the extensive margin). Note that interpretation of these estimates should incorporate the existence of consumption taxes, which vary around 20 percent in many European countries. Also, it is likely that the optimal MTRs are lower when the generosity of the income grant is set below the optimal level. Similarly, if work is deemed to have a social value then optimal benefit withdrawal will likely be lower. On the other hand, if conditionality is strong then optimal withdrawal rates may be higher, and conditionality may be less desirable when withdrawal rates are low.

18. This could be interpreted as the tax elasticity being much higher at the bottom of the income distribution reflecting significant evasion possibilities, thus making high tax rates at low incomes very inefficient. Note that information (or data) constraints are at the heart of the design problem, as well as constraints on delivering transfers, and these constraints are typically most challenging in countries with large informal sectors and weak administrative capacity (see Prady, 2020, for a discussion and recent initiatives). The strength of public financial management (PFM) systems can also drive political will to undertake fiscal redistribution.
19. Note that our discussion of empirical results based on descriptive regression analysis should be seen as simply a way of describing associations (or 'correlations') between variables rather than suggesting any underlying causal relationship.
20. The contribution of an explanatory variable (X) to the total variance of the dependent variable (Y) is calculated as the coefficient estimate times Cov(X,Y) /Var(Y). An alternative approach would be to use the Shapely-Shorrocks decomposition (see Coady et al., 2021; Shorrocks, 2013) for an example in the current context.

REFERENCES

Ahmad, Ehtisham, Nicholas Stern, 1991, *The Theory and Practice of Tax Reform in Developing Countries*, (Cambridge: Cambridge University Press).

Atkinson, Anthony, 1970, 'On the Measurement of Inequality,' *Journal of Economic Theory*, Vol. 2 (September), pp. 244–63.

_______, Joseph Stiglitz, 1980, *Lectures on Public Economics*, (London: McGraw-Hill).

_______, 1995a, 'On Targeting and Family Benefits,' in *Incomes and the Welfare State: Essays on Britain and Europe*, ed. by Anthony Barnes Atkinson (Cambridge: Cambridge University Press).

______, 1995b, 'On Targeting Social Security: Theory and Western Experience with Family Benefits,' in *Public Spending and the Poor: Theory and Evidence*, ed. by Dominique van de Walle and Kimberly Nead (Baltimore: Johns Hopkins University Press).

Atkinson, Anthony, and Andrea Brandolini, 2010, 'On Analyzing the World Distribution of Income,' *World Bank Economic Review*, Vol. 24, No. 1, pp. 1–37.

Avram, Silvia, Horacio Levy, and Holly Sutherland, 2014, 'Income Redistribution in the European Union,' *IZA Journal of European Labor Studies*, Vol. 3.

Bargain, Olivier, 2017, 'Welfare Analysis and Redistributive Policies,' *Journal of Economic Inequality*, Vol. 15, pp. 393–419.

Besley, Timothy, and Ravi Kanbur, 1993, 'The Principles of Targeting,' in *Including the Poor*, ed. by Michael Lipton and Jacques Van Der Gaag (Washington: The World Bank Group).

Brewer, Mike, Emmanuel Saez, and Andrew Shepard, 2010, 'Means-testing and Tax Rates on Earnings,' pp. 91 in *Dimensions of Tax Design*, ed. by Stuart Adam, Tim Besley, Richard Blundell, Stephen Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James M. Poterba (Oxford: Oxford University Press).

Caminada, Koen, Jinxian Wang, Kees Goudswaard, and Chen Wang, 2017, 'Income Inequality and Fiscal Redistribution in 47 LIS Countries, 1967-2014,' LIS Working Paper No. 724, (Luxembourg: Cross-national Data Centre).

_______, Kees Goudswaard, Chen Wang, and Jinxian Wang, 2019, 'Income Inequality and Fiscal Redistribution in 31 Countries After the Crisis,' *Comparative Economic Studies*, Vol. 61, pp. 119–148.

Causa, Orsetta, and Mikkel Hermansen, 2019, 'Income Redistribution Through Taxes and Transfers Across OECD Countries,' OECD Working Papers, No. 1453, (Paris: Organisation for Economic Co-operation and Development).

Chetty, Raj, 2006, 'A New Method for Estimating Risk Aversion,' *American Economic Review*, Vol. 96, No. 5, pp. 1821–1834.

Clements, Benedict, Ruud de Mooij, Sanjeev Gupta, Michael Keen, eds, 2015, *Inequality and Fiscal Policy*, (Washington: International Monetary Fund).

Coady, David, and Emmanuel Skoufias, 2004, 'On the Targeting and Redistributive Efficiencies of Alternative Transfer Programs,' *Review of Income and Wealth*, Vol. 50, pp. 11–27.

Coady, David, Margaret Grosh, and John Hoddinott, 2004a, 'Targeting of Transfers in Developing Countries: Review of Lessons and Experience,' World Bank Regional and Sectoral Studies No. 30230 (Washington: The World Bank Group).

_______, 2004b, ‘Targeting Outcomes Redux,’ *The World Bank Research Observer*, Vol. 19, No. 1, pp. 61–85 (Washington: The World Bank Group).

Coady, David, Devin D’Angelo, and Brooks Evans, 2021, ‘Fiscal Redistribution, Social Welfare and Income Inequality: “doing more” or “more to do”?’, *Applied Economics*, published online 18 Oct 2021.

Coady, David, Samir Jahan, Baoping Shang, and Riki Matsumoto, 2021, ‘Guaranteed Minimum Income Schemes in Europe: Landscape and Design,’ IMF Working Paper No. 2021/179 (Washington: International Monetary Fund).

Coady, David and Nghia-Piotr Le, 2020, ‘Designing Fiscal Redistribution: The Role of Universal and Targeted Transfers,’ IMF Working Paper No. 2020/105 (Washington: International Monetary Fund).

Crepaldi, Chiara, et al., 2017, ‘Minimum Income Policies in EU Member States,’ Directorate General for Internal Policies (Brussels: European Parliament).

Deaton, Angus, 1997, *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*, (Baltimore: John Hopkins University Press).

Deaton, A., and John Muellbauer, 1983, *Economics and Consumer Behavior*, (Cambridge University Press).

Diamond, Peter. A, 1975, ‘A Many-person Ramsey Tax Rule,’ *Journal of Public Economics*, Vol. 4, pp. 335–42.

Frazer, Hugh, Eric Marlier, 2016, ‘Minimum Income Schemes in Europe: A study of national policies 2015,’ Directorate General for Employment, Social Affairs and Inclusion (Brussels: European Commission).

Immervoll, Herwig, 2009, ‘Minimum-Income Benefits in OECD Countries: Policy Design, Effectiveness and Challenges,’ IZA Discussion Paper No. 4627.

_______, Linda Richardson, 2011, ‘Redistribution Policy and Inequality Reduction in OECD Countries: What Has Changed in Two Decades?,’ OECD Social, Employment and Migration Working Papers, No. 122 (Paris: Organisation for Economic Co-operation and Development).

Inchauste, Gabriela, and Nora Lustig, 2017, *The Distributional Impact of Taxes and Transfers: Evidence from Eight Low- and Middle-Income Countries*, (June) (Washington: World Bank).

Kanbur, Ravi, Michael Keen, and Miatti Tuomala, 1994, ‘Labor Supply and Targeting in Poverty Alleviation Programs,’ *World Bank Economic Review*, Vol. 8, pp. 191–211.

_______, Miatti Tuomala, 2011, ‘Charitable Conservatism, Poverty Radicalism and Inequality Aversion,’ *Journal of Economic Inequality*, Vol. 9, pp. 417–431.

Mirrlees, J.A. 1971, ‘An Exploration in the Theory of Optimum Income Taxation,’ *The Review of Economic Studies*, Vol. 38, No. 2, pp. 175-208

Newbery, David, and Nicholas Stern, 1987, *The Theory of Taxation for Developing Countries*, (New York: Oxford University Press).

Organisation for Economic Co-operation and Development (OECD), 1997, ‘Making Work Pay: Taxation, Benefits, Employment and Unemployment,’ (Paris: OECD Publications).

_______, 2008, ‘Growing Unequal: Income Distribution and Poverty in OECD Countries,’ (Paris: OECD Publications).

_______, 2011, ‘Divided We Stand, Why Inequality Remains Rising,’ (Paris: OECD Publications).

Pena-Cassas, Ramon, et al., 2013, ‘Towards a European Minimum Income,’ EESC Study No. EESC/COMM/03/2013 (Brussels: European Economic and Social Committee).

Piketty, Thomas, and Emmanuel Saez, 2013, ‘Chapter 7 - Optimal Labor Income Taxation,’ *Handbook of Public Economics*, Vol. 5, pp. 391–474.

Prady, Delphine, 2020, ‘Reaching Households in Emerging and Developing Countries: Citizen ID, Socioeconomic Data, and Digital Delivery,’ Special Series on Fiscal Policies to Respond to COVID-19 (Washington: International Monetary Fund).

Sen, Amartya, 1988, ‘Chapter 1: The Concept of Development,’ Handbook of Development Economics, Vol. 1, pp.10–25

Shorrocks, Anthony, 2013, ‘Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value,’ *Journal of Economic Inequality*, Vol. 11, pp. 99–126.

Sutherland, Holly, and Francesco Figari, 2013, ‘EUROMOD: The European Union Tax-Benefit Microsimulation Model,’ *International Journal of Microsimulation*, Vol. 6, pp. 4–26.

van de Walle, Dominique, 1998, ‘Targeting Revisited,’ *The World Bank Research Observer*, Vol. 13, No. 2, pp. 231–48 (Washington: The World Bank Group).

15. Efficiency of public investment and implications for the optimal level of public investment

Jean-Marc Fournier and Fabien Gonguet[1]

1. INTRODUCTION

Public investment (in)efficiency has been attracting increasing attention in the public investment literature since seminal works of the World Bank (1994), Hulten (1996) and Pritchett (2000). Public officials are the agents of the people who are the principal, and their interests are not fully aligned, potentially leading to sub-optimal choices. Pritchett found with cross-country data that one dollar of public investment translates into about 50 to 60 cents of public capital, and that inefficiency is substantially lower in OECD and fast-growing East Asia countries than in most other developing economies.

Despite these early warnings, the concerns remain sizeable and recent estimates point to the existence of significant margins to improve public investment efficiency in most countries. Miyamoto et al. (2020) showed that emerging markets and low-income developing countries still trail significantly behind advanced economies in terms of infrastructure access and quality, despite sustained larger public investment to GDP ratios over the past three decades. Baum et al. (2020) found an average efficiency gap relative to best performers of one-third to two-fifths (depending on the estimation methodology) in a study covering 164 countries, with smaller gaps as income rises. Low-income developing countries lose on average 44 to 53 percent of their investment resources to inefficiencies, against 32 to 42 percent in emerging markets and 15 to 27 percent in advanced economies.

These concerns for public investment efficiency are particularly acute as public capital is a driver of economic growth, as shown by a voluminous literature. Earlier research has embedded public investment in growth theory to discuss the optimal level of public investment (Arrow and Kurz 1970). As gains from some public infrastructures are not excludable, the private market would fail to provide the necessary services, so that public capital is likely to crowd in private capital (for example Aschauer 1989). However, public investment should not displace better opportunities in the private sector, in which case private investment would be crowded out (Marglin 1963). This suggests that the effect of public investment is an empirical question, and a first generation of empirical research showed large effects of public investment on growth (Aschauer 1989; Fernald 1999). This has been followed by a voluminous literature providing heterogeneous results that are well summarized by Bom and Ligthart's (2014) meta-analysis. Overall, they find positive but more modest effects of public investment, in part because the econometric specification in Aschauer (1989) has been improved to address the presence of stochastic trends and of reverse causality. In the wake of the COVID-19 pandemic, infrastructure investment was highlighted as a key ingredient for a sound economic recovery and for the transition to a more inclusive and greener economic model, while insisting on the need to enhance its efficiency and quality to achieve higher benefits (IMF 2020).

Given the large and persistent concerns for public investment inefficiency, this chapter aims at providing a concise overview to help researchers and policy makers understand the issue and think about appropriate policy responses. It discusses how to establish the right diagnostic, and how policies could improve the gains reaped from public investment. There has been a growing literature on measuring public investment efficiency and on the positive impact of sound public investment management (PIM) practices on efficiency. The literature on the appropriate depth of public capital deepening given these concerns is much scarcer. This chapter will thus review the literature on how to measure and improve public investment efficiency and complement this with a distinct and original analysis of the effect of public spending inefficiency on the optimal rate of investment. This shows that utility-maximizing public investment level increases with efficiency when public investment substitutes private factors of production, while it decreases with efficiency otherwise.

The remainder of the chapter is organized as follows. Section 2 briefly defines public investment efficiency and section 3 describes how it can be measured. Section 4 discusses how public investment efficiency can be improved. Section 5 provides insights into the right amount of public investment, being cognizant of public investment inefficiency.

2. DEFINING PUBLIC INVESTMENT EFFICIENCY

A common definition of public investment efficiency is the fraction of investment that passes into the value of capital (Hulten 1996, Pritchett 2000). Formally, this consists in adding an efficiency factor in the public capital accumulation equation:

$$K_t = (1 - \delta)K_{t-1} + \varepsilon_t I_t \tag{15.1}$$

where K_t is public capital, δ is the rate of depreciation of public capital, ε_t is public investment efficiency and I_t is public investment. While I_t is the amount spent, $\varepsilon_t I_t$ is the economic cost of public investment, namely the minimum achievable cost for the same investment. Usual models that assume full efficiency thus overestimate the capital stock, so that the decomposition of output into the contribution of capital and total factor productivity is incorrect (Pritchett 2000).[2]

As discussed in Pritchett (2000), the value of public capital can differ from investment cost not only because of waste or mistakes, but also because technological change can make old investment obsolete, and because relative prices differ over time or across countries, including due to tariffs. Private sector capital stock can be affected by technological surprises as well as the public sector, and private investors are expected to maximize the value of their investment *ex ante*, given their own knowledge and technology when they make their decisions. Differences in prices across countries can alter cross-country comparisons in nominal spending (De Long and Summers 1992), this can be fixed with investment series based on comparable prices of investment goods across countries (as in King and Levine 1994). A key difference between public and private investment is that private decision makers are incentivized to minimize waste so as to maximize their profits and those who make inefficient investment decisions may have to exit markets, while there is no guarantee *a priori* that public decision makers would minimize mistakes or waste. To keep focus on the latter, which is inherent to the public sector, the definition of public investment efficiency shall steer clear from changes in technologies or prices that were not predictable.

With this definition, causes of public investment inefficiency include waste, poor project selection, insufficient maintenance and corruption. Examples of wasteful projects abound (for example Schwartz et al. 2020). There is waste if poorly planned projects are not completed, when there is useless bureaucratic work or if implementation is less productive because it does not use available technology. According to the IMF (2020), almost 40 percent of projects cost more than the estimated appraisal cost and 75 percent of projects are delayed beyond their projected completion date at project outset. Poor appraisal may distort project selections. Further, this selection may also be distorted by political considerations, such as EU structural fund allocation favoring regions which are politically aligned with national authorities (Dotti 2015). As new projects attract visibility, maintenance is often structurally underfunded, accelerating the depreciation of existing projects and increasing fiscal costs in the medium to long run (Blazey et al. 2020). Insufficient maintenance has been identified early as a source of unnecessary costs. For instance, the World Bank flagged in 1994 that $12 billion in timely road maintenance in Africa would have avoided $45 billion in rehabilitation (World Bank 1994). Finally, relative to other types of spending, investment projects typically have unique features, increasing the level of discretion, complicating cost comparisons and thus making it easier to conceal bribes (Mauro 1998; IMF 2019), so that infrastructure costs are higher in countries with higher corruption (Collier et al. 2016). Complexity also reduces capacity to control efficiency of public officials in implementing projects.

Waste, poor project selection and corruption can be directly associated with this definition of efficiency as discussed in Berg et al. (2019). First, by definition, waste implies that one unit of investment does not translate into one unit of public capital. Second, a unit of investment in a poorly chosen project delivers fewer public services than a unit of investment in the optimal project. Third, corruption is also obviously consistent with this capital accumulation equation, but this does not capture the full picture as corruption is also a transfer toward a corrupt official that would consume this benefit, as discussed in Berg et al. (2019). Last, one may consider that lack of maintenance is a form of poor project selection, this thus can fit with the same inefficiency definition. However, one could also reflect this with a higher depreciation rate. With this alternative approach, the efficiency parameter ε_t would not capture the role of poor maintenance. If the depreciation rate is not allowed to vary with maintenance quality, as discussed below, all sources of inefficiency are embedded in ε_t.

3. MEASURING PUBLIC INVESTMENT EFFICIENCY

Measures of public investment efficiency compare production units (typically the countries), assuming those that get the largest outputs (for example kilometers of roads, perception of quality of infrastructure and so on) for a given amount of inputs, including cumulated past investment spending, are efficient (see Murillo-Zamorano 2004 for a survey). These units are at the efficiency frontier and, for other units, the distance to this frontier is a measure of inefficiency. Parametric and non-parametric methods are used in the literature. The main parametric approach is stochastic frontier analysis (SFA), in which the deviation from a parametric production function is made up of two terms – statistical noise and inefficiency (see Kumbhakar et al. 2020 for a recent overview). The aim of the statistical noise is to capture factors such as luck or uncertainties that should not belong to efficiency. Among the non-parametric approaches, data envelopment analysis (DEA) is widely used (for example IMF 2015; Schwartz et al. 2020). In this approach, no functional form for production is assumed, the frontier is the piecewise linear envelope that encompasses all units and inefficiency is the distance to this frontier. A noticeable improvement from the seminal analysis by Charnes et al. (1978) has been to allow variable returns to scale that enables each unit to be compared to other similar alternatives (Byrnes et al. 1984).

Implementing such efficiency measures requires calculating inputs and outputs. As the output is a stock of infrastructure or the service they provide, practitioners consider public capital stock rather than public investment. The input measure reflects the amount spent, such as a public capital stock calculated with the perpetual inventory method. Measures that favor comparability rather than accuracy with a common depreciation rate across countries, as in Kamps (2006) – which has been expanded and updated in IMF (2017) – are relevant. Higher depreciation in some countries for a given sector typically reflects lower investment quality, one does not want to adjust the input measure downward before computing an efficiency score. As full efficiency is achieved when the government is getting the most out of public investment with the available technology, Baum et al. (2020) also use GDP per capita as a complementary input, aiming at ruling out the role of lower technology in low-income countries.

As pertains outputs, since the seminal work of Canning (1998), researchers have been gathering data on infrastructure such as road length, number of telephone lines or energy production, for instance, as in Donaubauer et al. (2016). These databases may be further improved with new internet data such as online map routine services, as illustrated by Braconier and Pisu (2013), who compare road distance to the great circle distance. The contribution of the private sector to physical infrastructure outputs can create a scope mismatch in efficiency estimates, which can be mitigated by restricting the analysis to periods when the share of the private sector in infrastructure investment is low, as in Hurlin and Arestoff (2010). Further, heterogeneity of sectoral composition of public investment may need to be considered. Despite all the caveats, public capital stock outputs need to rely on such measures because there is no market to gauge the value of public capital in those many cases in which public capital is a public good that is not traded.

As public capital accumulation takes time, such measures reflect moving averages of public efficiency. For instance, with a growth rate of 2 percent and a depreciation rate of 4 percent, the average age of public capital stock is about 16 years in the steady state and it is common that the life span of economic and social infrastructure exceeds several decades. As a result, there is a lack of analysis of the evolution of public investment efficiency over time in the literature. However, an alternative that would focus on public investment directly is unlikely to address this issue. It could miss the cost of accelerated depreciation due to poor construction quality for instance. In contrast, the whole capital stock encompasses all public investment issues. As time passes, there may be scope for an analysis over a very long period that would enable capturing changes in efficiency, this would require careful data work, especially to ensure consistency of output measures over time.

4. PUBLIC INVESTMENT MANAGEMENT (PIM) REFORMS TO IMPROVE PUBLIC INVESTMENT EFFICIENCY

4.1 PIM Frameworks

The wide dispersion of public investment efficiency across countries can be related to a large heterogeneity in policies. The rate of return of public investment can be shaped both by policies that affect investment in the whole economy and by policies that pertain to public investment more specifically. When the exchange rate, trade, pricing and regulatory policies are inappropriate, private and public investments are affected, as discussed by Isham and Kaufmann (1999). In contrast, PIM frameworks and practices affect public investment directly. To keep the focus on these sources of public investment efficiency for a given level of technology and macroeconomic environment, this section will focus on PIM.

There is ample empirical evidence (Gupta et al. 2014; IMF 2015; Miyamoto et al. 2020) that by improving institutions' ability to manage public investment, governments are likely to reinforce the link between public investment and growth. Baum et al. (2020) show a positive correlation between public investment efficiency and the overall soundness of PIM practices: an average country could close more than half its efficiency gap were it to align its PIM practices with best performers. As they are using an efficiency measure which controls for the level of development (GDP per capita), this link between efficiency and PIM goes beyond the fact that both can improve with the development level: while the inefficiency gap is wider on average for lower-income countries, better PIM practices are likely to increase efficiency whatever the income level.

Reflecting a rising attention to public investment efficiency to enhance the growth dividends of public infrastructure, international organizations have designed frameworks to help governments assess the strengths and weaknesses in their PIM practices and identify reform priorities. Dabla-Norris et al. (2012) introduced a Public Investment Management Index (PIMI), gauging policies at the appraisal, selection, implementation and evaluation phases, focusing on low-income and emerging markets. In the same vein, the World Bank, the International Monetary Fund (IMF) and the Organisation for Economic Co-operation and Development (OECD) have established multi-dimensional indicators to identify strengths and weaknesses in PIM policies (Rajaram et al. 2014; OECD 2017 and IMF 2018). The IMF Public Investment Management Assessment (PIMA) framework approaches infrastructure governance in a holistic manner by evaluating 15 different practices across all phases of the project cycle (IMF 2018; Chaponda et al. 2020). The holistic view is critical to sound advice on reform priorities, as there are links between several of these 15 practices (for example, project appraisal and project selection; availability of funding and monitoring of public assets; management of project implementation and monitoring of project portfolio). While PIM practices should be anchored by a strong legal and regulatory framework, the PIMA framework insists on the importance of looking beyond the institutional design (that is, what exists on paper) and considering the implementation (that is, what happens in practice) to understand where key reform priorities lie. Finally, Kim et al. (2020) aim at complementing these with the *Public Investment Management Reference Guide* with a particular focus on how to adopt the implied reforms, noting the demand for practical guidance.

4.2 PIM Reforms: Priorities and Challenges

These multi-dimensional tools revolve around the investment project cycle, which includes the planning, allocation and implementation phases; and they also consider cross-cutting enabling factors. In the planning phase, emphasis is placed on the need for investment plans across the public sector (central and subnational governments, state-owned enterprises) to be affordable and consistent with the government's overall development and sectoral strategies, and for projects to be carefully appraised. Sound allocation of budget resources to the best projects hinges on solid medium-term budget projections, coherence between current and capital budgets, consideration given to maintenance and the objective selection of projects for budget funding. The implementation phase includes the procurement of projects, the oversight of project delivery, mechanisms to ensure availability of funding and the monitoring of existing infrastructure assets. Cross-cutting enabling factors include, for instance, staff capacity and information technology systems.

The IMF PIMA has been applied to more than 70 countries so far. PIMA scores reveal that there is scope for all countries to enhance their PIM practices, bringing prospects of improved public investment efficiency (Chaponda et al. 2020). Average PIMA scores, detailed for each of the 15 practices, also provide an overview of where PIM weaknesses lie in a typical country. Investment planning is often stronger than allocation and project implementation. Appraisal and selection processes are frequently among the weakest links. Appraisals often suffer from lack of resources and limited skills; but even when resources and skills are adequate, it is common that insufficient time is devoted to appraisals, so that projects can be procured and delivered sooner. Weak project appraisal affects the quality of information used to select projects for inclusion in the budget. And when appraisal information is of high quality, there might still be political interference at the selection stage, especially when selection criteria are not transparently communicated. Adverse incentives can also play a role in the selection process as the best project on paper can also be the one with higher cost or time overruns (Flyvbjerg 2009). Another common weakness lies in the monitoring of public assets – only a small number of countries keep an up-to-date account of their infrastructure assets and their condition, thus preventing the sound calculation of maintenance needs.

Reforms in these areas – for example, shared appraisal guidelines and methodologies, funding and capacity development for appraisal, published selection criteria, asset registries – can be game changers in avoiding bad projects and in ensuring prolonged life spans of infrastructure. But while PIM reforms can significantly contribute to improving the efficiency of public investment, governments are often faced with several challenges and shortcomings which tend to slow down reform implementation.

Turning sound PIM principles into effects is often difficult. PIMA scores reveal that the design of PIM frameworks as set on paper is almost systematically stronger than what happens in practice. This gap between institutional design and effectiveness is an important element to consider in a PIM reform strategy. For instance, public procurement is, on average, an area where the principles of openness, fairness and transparency are relatively well reflected in the legal and regulatory framework. However, this is the PIMA practice for which the gap between institutional design and effectiveness is the largest (Chaponda et al. 2020). Despite good intentions in the legal framework for public procurement, single bidding, direct procurement modalities or unsolicited bids might still be prevalent. Multiyear budgeting is another example of an area where effectiveness can be lagging behind even when institutional design is appropriate. While changing a law or a regulation can be done relatively quickly, they are usually not met with immediate effects and often require awareness-raising efforts and capacity building over several years.

Coordination among stakeholders is a key ingredient for the successful delivery of PIM reforms, but it is difficult to achieve. Public investment projects can involve different ministries, different layers of administration and different public sector entities; and the diversity of financing arrangements (budget, external donor funding, public–private partnership) adds to the complexity. This creates a coordination challenge. As a high share of public investment occurs at the local level, there is a need to foster and improve policy coordination, transparency and information sharing across levels of government (Allain-Dupré 2011). For broad PFM institutions related to the budget process or to the financial oversight of other public sector entities, the Ministry of Finance is usually the lead agency; while for more specific or technical areas directly related to project design and management, line ministries and sector regulators are on the frontline. For reforms to be successful, a high-level commitment to reform from the ministry of finance is critical; but so is buy-in by other stakeholders.

Integration of planning and budgeting functions is one of the major challenges of PIM. Instruments that can support it, such as medium-term budget frameworks or public investment programs, have been advised and set up on paper, but have displayed relatively weak effectiveness on average (Chaponda et al. 2020). Political economy also plays a part in the integration (or lack thereof). In a majority of countries, planning and budgeting are undertaken by separate entities, typically the Ministry of Planning and the Ministry of Finance, respectively; a single central ministry is responsible for both planning and budgeting in 37 percent of countries (Allen et al. 2020). Having separate entities is not necessarily less efficient, as evidenced by the successful examples of Colombia or Ireland. Further, tensions can be just as strong between departments of a single ministry. But these institutional arrangements tend to change often over time, calling for new processes, with an uncertain effect on efficiency and effectiveness of investment choices.

At all stages of the public investment process, the government needs a sound legal and regulatory framework, reliable information technology (IT) systems and an adequate number of trained staff. The legal and regulatory framework should ensure clear mandates, transparent criteria, accountability and effective standards and procedures (IMF 2018). As pointed out by the OECD (2017) and Pattanayak and Verdugo-Yepes (2020), corruption threats need to be mitigated at all stages of the process, including with a risk-based approach and a clear delineation of responsibilities. The IT system needs to be comprehensive and integrated to best assist decision making and monitoring. Finally, adequacy of staff capacity (numbers and skills) is a pre-requisite to implement the principles discussed here. These factors can be critical bottlenecks in the success of PIM reforms, especially in lower-income countries. Capacity development by institutional partners such as the World Bank or the IMF is helpful, but usually occurs across several years.

In the context of the response to a crisis such as the COVID-19 pandemic, a timely and effective push to invest is recommended, especially in the light of higher fiscal multipliers in times of high uncertainty (IMF 2020). But PIM shortcomings, as well as the lack of trained staff and reliable IT systems, can hamper absorptive capacity, thus limiting the economic effects of a scaling up of investment. Presbitero (2016) indeed shows that, when the public investment to GDP ratio is high compared to the past five-year average, the share of project rated as satisfactory is lower, although that effect is small. Gurara et al. (2020) find that this concern is sharper in low-efficiency countries, where unit costs increase once public investment is close to about 7 percent of GDP, against 10 percent of GDP on average. Melina et al. (2016) provide simulations illustrating the merit of smoothing public investment in the presence of such absorptive capacity constraints.

It is still possible to ramp up investment in the short run to support the recovery, as recommended by the IMF (2020), provided there is some degree of good PIM practice. For example, governments can consider focusing on infrastructure maintenance projects, which are of smaller size, more quickly delivered and of a limited degree of complexity. Maintenance backlogs are large – for example 3.5 percent of GDP (one-off expenditure) for highways and bridges in need of repair in the United States (ASCE 2018). The efficiency of this short-term strategy still hinges on the ability of governments to identify where the most pressing needs are. To do so, an asset registry is required, or, at least, field knowledge of sectoral ministries or subnational governments. Another short-term action is to review and, if needed, reprioritize ongoing projects, so as to leave fiscal space for new priorities and ensure that projects still respond to current needs and their projected financials still rely on realistic assumptions, in spite of the crisis. This requires the ability to actively monitor investment projects. Overall, while quick fixes or shortcuts are always possible, the recovery timeframe is too short to implement a full-fledged PIM reform strategy that would really allow governments to reap the full investment efficiency potential.

5. INEFFICIENCY AND THE OPTIMAL SHARE OF PUBLIC INVESTMENT

Given the sizeable concerns about public investment inefficiency, this section explores the extent to which this should lead policy makers to adjust the amount of public investment. While they should be an area of priority for governments to unlock the potential of their investment, structural reforms to improve PIM processes as discussed in section 4 take time, and the speed and size of efficiency gains reaped from reforms is uncertain. It can thus be realistic and cautious to take inefficiency as given when thinking of optimal public investment decisions in the short to medium run.[3]

Despite the growing literature on public investment efficiency, research on the appropriate implications for optimal public investment size is scarce. Agénor (2010) assumes network effects with which public investment efficiency can increase beyond a public capital stock threshold, leading to multiple equilibria, so that a big public investment push can enable a country to switch to a higher equilibrium. However, such higher returns may not hold in the presence of congestion (Aschauer 1989), and indeed Isham and Kaufmann (1999) and Fournier (2016) rather report decreasing marginal returns of public investment. Van der Ploeg (2012) discusses the implication of windfall revenues in the presence of public investment inefficiency. Berg et al. (2013) run simulations in the presence of public investment inefficiency in resource-rich developing countries to illustrate that the public investment level needs to factor in future maintenance costs which can become excessive when windfall revenues stop. More generally, Berg et al. (2019) show that, with a Cobb-Douglas production function, the effect of public investment on growth does not depend on efficiency.[4] Lower efficiency not only reduces the increase in public capital stock for a unit of investment, but also increases the marginal gain from this additional unit as public capital stock is lower, and these two mechanisms offset each other.

However, these papers do not discuss the effect of public investment inefficiency on optimal public investment levels, and this chapter aims at filling in this gap. The public investment to output ratio maximizing consumption in a simple golden-age growth path (Allais 1947[5]) is formulated as a function of public investment efficiency.

5.1 Implication of Inefficiency for the Optimal Investment Rate on a Golden-Age Growth Path

Output is produced with a constant return to scale production function with an exogenous technological progress $e^{\lambda t}$ making use of two factors of production; labor growing at an exogenous rate γ that is $L(t) = e^{\gamma t} L(0)$, and public capital $K(t)$

$$Y(t) = F(K(t), e^{\lambda t} L(t)) \tag{15.2}$$

$k(t)$ denotes capital per unit of effective labor $K(t)/e^{(\lambda+\gamma)t} L(t)$, so that the production function can be rewritten:

$$Y(t) = e^{(\lambda+\gamma)t} f(k(t)) \tag{15.3}$$

where

$$f(k(t)) = F(k(t), 1) \tag{15.4}$$

This analysis which differs from the golden rule of accumulation of capital model presented in Phelps (1965) is one aspect only, which is the addition of an efficiency parameter as in Pritchett (2000):

$$\dot{K}(t) = \varepsilon I(t) - \delta K(t) \tag{15.5}$$

As in Phelps (1965), there is a steady growth path with growth rate $g=\lambda+\gamma$

$$\dot{K}(t) = (1+g)K(t) \tag{15.6}$$

Hence at the steady state public investment shall compensate capital depreciation and enable a rise of capital stock at the same pace as output, and to achieve this investment needs to be rescaled to offset inefficiency losses:

$$I(t) = \frac{\delta+g}{\varepsilon} K(t) \tag{15.7}$$

When efficiency is not observed, public capital stock is measured with the inventory method assuming efficiency is one:

$$I(t) = (\delta + g)K^*(t) \tag{15.8}$$

The people consume the production which is not invested:

$$C(t) = Y(t) - I(t) \geq 0 \tag{15.9}$$

One can thus recalculate the optimal saving rate of the golden rule with capital efficiency losses. If $f'(0) > \frac{\delta+g}{\varepsilon}$ and $\lim_{k\to\infty} f'(k) < \frac{\delta+g}{\varepsilon}$, there is an interior solution:

$$\frac{\partial C(t)}{\partial K(0)} = \frac{\partial Y(t)}{\partial K(0)} - \frac{\partial I(t)}{\partial K(0)} \tag{15.10}$$

$$0 = f'(k) - \frac{\delta+g}{\varepsilon} \tag{15.11}$$

This provides the optimal capital per unit of effective worker ratio:

$$k(t) = f'^{-1}\left(\frac{\delta+g}{\varepsilon}\right) \tag{15.12}$$

Investment needed to sustain this capital level is thus calculated with equation (15.7)

$$I(t) = \underbrace{\frac{\delta+g}{\varepsilon}}_{(i\ to\ reach\ k)} \underbrace{f'^{-1}\left(\frac{\delta+g}{\varepsilon}\right)e^{(\lambda+\gamma)t}L(0)}_{(k\ target)} \tag{15.13}$$

which can be reformulated as a steady state investment to output ratio

$$i = \frac{I(t)}{Y(t)} = \underbrace{\frac{\delta+g}{\varepsilon}}_{i\ to\ reach\ k} \underbrace{\frac{f'^{-1}\left(\frac{\delta+g}{\varepsilon}\right)}{f\left(f'^{-1}\left(\frac{\delta+g}{\varepsilon}\right)\right)}}_{(k\ target)} \tag{15.14}$$

Two offsetting forces are shaping the optimal ratio of public investment to GDP. On one side, to achieve a given capital target one needs more public investment to offset losses if efficiency is lower. At the same time, the lower public investment efficiency, the lower the capital target should be.

A constant elasticity of substitution (CES) production function enables us to compute the condition on the elasticity of substitution for optimal public investment to increase with public investment efficiency:

$$f(k) = \left(ak^{\frac{\theta-1}{\theta}} + 1 - a\right)^{\frac{\theta}{\theta-1}} \tag{15.15}$$

This specification allows convenient simplifications:

$$f'(k) = a\left(a + (1-a)k^{\frac{1-\theta}{\theta}}\right)^{\frac{1}{\theta-1}} \tag{15.16}$$

$$f'^{-1}(k) = \left(\frac{\left(\frac{k}{a}\right)^{\theta-1} - a}{1-a}\right)^{\frac{\theta}{1-\theta}} \tag{15.17}$$

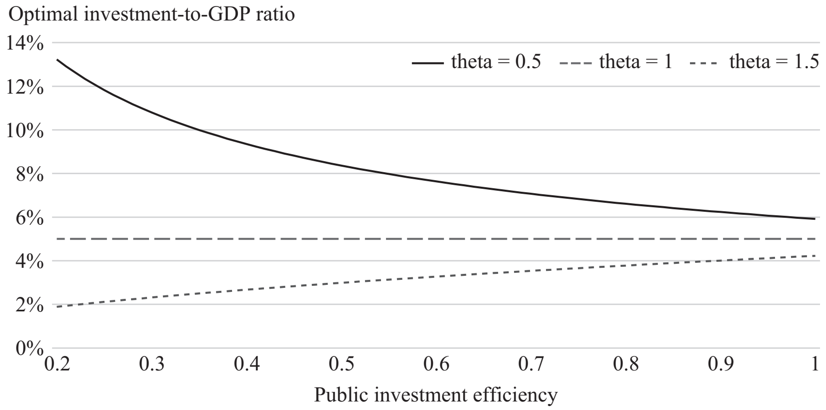
$$\frac{f'^{-1}(k)}{f(f'^{-1}(k))} = \left(\frac{k}{a}\right)^{-\theta} \tag{15.18}$$

The investment rate is a function of the public capital share in the production function a, the elasticity of substitution θ, capital depreciation rate δ, growth rate g and public investment efficiency ε, which is the interior solution if $\theta \geq 1$ or $\varepsilon > (\delta + g)/a^{\theta/(\theta-1)}$:[6]

$$i = \underbrace{\frac{\delta+g}{\varepsilon}}_{(i \text{ to reach } k)} \underbrace{\left(\frac{a\varepsilon}{\delta+g}\right)^{\theta}}_{(k \text{ target})} = a^{\theta}\left(\frac{\delta+g}{\varepsilon}\right)^{1-\theta} \tag{15.19}$$

The effect of public investment efficiency on the optimal public investment to GDP ratio crucially hinges on the complementarity of public capital (equation 15.19 and Figure 15.1). In the intermediate case of a Cobb-Douglas production function with $\theta = 1$, this optimal ratio is the share of public capital in the production function a. It does not depend on public investment efficiency. When efficiency decreases, for a given capital target, the government would need to invest more to compensate waste. This is offset by the need to target a lower capital stock. This echoes the efficiency paradox identified by Berg et al. (2019), namely, that the effect of public investment on growth can increase when efficiency decreases if the elasticity of substitution is low. The Cobb-Douglas case delineates the higher complementarity case ($\theta < 1$) in which the lower public investment efficiency, the higher public investment should be and the higher substitutability case ($\theta > 1$) in which the opposite conclusion holds. A similar finding holds if one considers the depreciation rate δ as a measure of public investment inefficiency. In the Cobb-Douglas case, the optimal investment ratio does not depend on the depreciation rate, and when the elasticity of substitution is high, then a higher depreciation rate implies a lower optimal public investment ratio.

This result should be read in terms of complementarity between public capital and private production factors which are illustrated here with labor – the model does not include private capital for a simple and intuitive exposure. In the same vein, if public investment is highly complementary to private capital – like building the roads or railways that connect plants to a port, then public investment catalyzes private investment. Thus, when efficiency is low and complementarity is high, more needs to be spent to provide those much needed public



Note: Theta is the elasticity of substitution between public capital and labor. Steady-state consumption-maximizing public investment ratio with share of public capital a = 5 percent, depreciation rate of public capital δ = 5 percent and growth rate g = 2 percent.

Figure 15.1 Optimal public investment-to-GDP ratio, efficiency and public capital complementarity

infrastructures.[7] Berg et al. (2019) consider a nested production function with an elasticity of substitution between public and private capital and an elasticity of substitution between capital and labor and provide conditions on elasticities of substitutions for the efficiency paradox to hold, including when private capital is reacting endogenously and during a transition path.[8] This confirms the finding that complementarity between public capital and private factors of production generates this paradox.

The importance of substitutability is intuitive. The high complementarity case applies to public investment projects which, for certain reasons, are necessary to enable private sector activity, such as the provision of roads to connect industrial plants and people, without which these cannot work. Then, even if efficiency is low, the capital target should not be reduced much as this is a critical input, and more investment is needed as part of it is wasted. In contrast, if the government is providing electricity while private firms can provide alternative sources of energy, then the investment is worth it only if it is efficient. Otherwise, better if this is done by the private sector.

The effect of public investment on growth in a country which implements the optimal public investment ratio can be written as follows:[9]

$$\frac{\partial Y}{\partial I}\frac{I}{Y}=\left(\frac{\partial Y}{\partial k}\frac{\partial k}{\partial I}\right)\frac{I}{Y}=e^{(\lambda+\gamma)t}f'(k)\cdot\frac{\varepsilon}{e^{(\lambda+\gamma)t}L}i \tag{15.20}$$

Equation (15.11) provides *f'(k),* such that:

$$\frac{\partial Y}{\partial I}\frac{I}{Y}=\frac{\delta+g}{L}i=\frac{\delta+g}{L}a^{\theta}\left(\frac{\delta+g}{\varepsilon}\right)^{1-\theta} \tag{15.21}$$

This shows that the role of efficiency in the growth effect of public investment is proportional to the role of efficiency in the optimal public investment rate. The effect of public investment on growth thus increases with efficiency when public investment is a substitute for the private factor of production, while it is the opposite when it is a complement. The role of efficiency in shaping effects of public investment on growth is thus an empirical question.

Two empirical approaches lead to consistent conclusions that, in most countries, public capital is a substitute to private factors of production. A first approach is to estimate elasticities of substitution. An et al. (2019) find an elasticity of substitution between public and private capital of around 3.[10] At the same time, a Gechert et al. (2019) meta-analysis and recent estimates by Chirinko and Mallick (2017) suggest that the elasticity of substitution between capital and labor is likely to be close to 0.4. Numerical results from Berg et al. (2019) show that with such an elasticity of substitution between capital and labor, the efficiency paradox does not hold for an elasticity of substitution between public and private capital above 1.4. All this suggests that higher efficiency should lead to a higher effect of public investment on growth in most cases. Further, such estimates may be refined to reflect country-specific production features or levels of development. A second approach is to explore the effect of growth. Abiad et al. (2016), Furceri and Li (2017) and Miyamoto et al. (2020) find that the growth effect of public investment increases with efficiency, a result that is consistent with estimates of elasticities of substitution.

Beyond these aggregate results, complementarity of capital may be higher in some areas. In particular, Canning and Bennathan (2000) find electricity capacity and road infrastructure complementarity with both physical and human capital, suggesting that, in these two areas, public investment may need to be higher to reach capital stock targets in the presence of inefficiencies. Also, Arezki et al. (2017) report that the global privatization experiment of the past three decades may have held back the supply of large-scale infrastructure projects, suggesting that, in some areas, the private sector was a poor substitute.

5.2 Deadweight Taxation Cost, Inefficiency and Optimal Public Investment

As taxation distorts economic choices, there can be a deadweight cost of taxation used to finance public investment which is assumed to be equivalent to a penalty $g(I(t))$ increasing in investment on the production function:

$$Y(t) = F(K(t), e^{\lambda t} L(t)) - g(I(t)) \tag{15.22}$$

Without specifying the function g, this formulation can embed any form of taxation which is realistic given the complexity of tax schemes in practice. As public investment is only a fraction of public expenditure, the deadweight loss of taxation may be linearized as follows:

$$Y(t) = F(K(t), e^{\lambda t} L(t)) - g_0 - g_1 I(t) = F(K(t), e^{\lambda t} L(t)) - g_0 - g_1 \frac{\delta + g}{\varepsilon} K(t) \tag{15.23}$$

Assuming that governments implement the least distortive taxes first, the function g shall be convex, so that with $g(0) = 0$, g_0 is negative. The optimization problem is similar as the one solved above after replacing the function f by $\tilde{f}$:

$$\tilde{f}(k) = f(k) - g_0 - g_1 \frac{\delta + g}{\varepsilon} k \tag{15.24}$$

$$\tilde{f}'^{-1}(k) = f'^{-1}\left(k + g_1 \frac{\delta + g}{\varepsilon}\right) \tag{15.25}$$

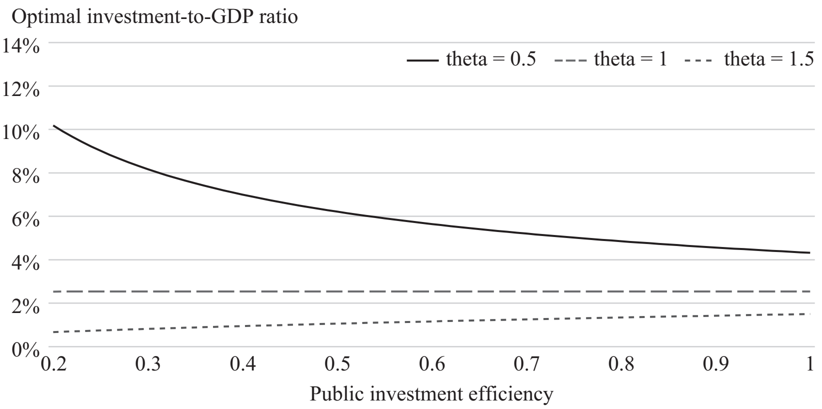
So that the optimal public investment ratio is

$$i = \frac{I(t)}{Y(t)} = \underbrace{\frac{\delta + g}{\varepsilon}}_{i\ to\ reach\ k} \frac{f'^{-1}\left((1+g_1)\frac{\delta + g}{\varepsilon}\right)}{\underbrace{f\left(f'^{-1}\left((1+g_1)\frac{\delta + g}{\varepsilon}\right)\right) - g_0 - g_1 \frac{\delta + g}{\varepsilon} f'^{-1}\left((1+g_1)\frac{\delta + g}{\varepsilon}\right)}_{k\ target}} \tag{15.26}$$

With a CES production function, this can be reformulated:

$$i = \frac{I(t)}{Y(t)} = \frac{(1+g_1)^{-\theta} a^{\theta} \left(\frac{\delta + g}{\varepsilon}\right)^{1-\theta}}{1 - \frac{g_0}{f\left(f'^{-1}\left((1+g_1)\frac{\delta + g}{\varepsilon}\right)\right)} - g_1 (1+g_1)^{-\theta} a^{\theta} \left(\frac{\delta + g}{\varepsilon}\right)^{1-\theta}} \tag{15.27}$$

In practice, while taxation deadweight cost can have an effect on the optimal level of public investment, its effect on the sensitivity of optimal investment with respect to efficiency is a second order issue. This is illustrated with Figure 15.2, with an arbitrarily high marginal (respectively average) deadweight cost of 1 percent of public investment equivalent to 1 percent (respectively 0.6 percent in the Cobb-Douglas case). This calibration implies a large



Note: Theta is the elasticity of substitution between public capital and labor. Steady-state consumption-maximizing public investment ratio with share of public capital $a = 5$ percent, depreciation rate of public capital $\delta = 5$ percent, growth rate $g = 2$ percent and a marginal deadweight cost associated with one percent of public investment equal to one percent of GDP, and an average cost of 0.005 percent.

Figure 15.2 Optimal public investment-to-GDP ratio, efficiency and public capital complementarity in the presence of deadweight taxation reduction of the optimal public investment ratio to reduce the deadweight tax cost. Even so, the effect of efficiency on the optimal public-investment-to-GDP ratio is almost not affected by deadweight taxation. In particular, in the Cobb-Douglas case, this optimal public investment ratio is almost independent from public investment inefficiency. This is because the marginal cost of taxation is not very different in a high-efficiency country and in a low-efficiency country.

6. CONCLUSION

In sum, there is a strong case for 'investing in investing' as Collier (2007) says. Governments should devote time, energy and money to understanding the causes of investment inefficiency and ways to provide remedies. Improving efficiency has an unambiguous direct effect on output and welfare.

However, this is easier said than done, and public investment inefficiency remains pervasive. Even if many efforts have been made, issues such as weak project appraisal, insufficient maintenance, inadequate project selection or corruption remain topical today. Improving efficiency can only be achieved with deep changes that require time, government capacity and strong commitment to reform.

Being cognizant of the difficulty in reaching the first best of full efficiency, policy makers can also be guided by a second best – what is the appropriate level of public investment in the presence of inefficiency? The analysis of this chapter shows that the implication of inefficiency for the optimal public investment to GDP ratio crucially hinges on complementarity with private factors of production. The most common case is likely to be one with enough substitutability for an optimal public investment to GDP ratio to increase with efficiency. But when a public investment project is highly complementary to private factors of production, it can be worth spending more public money to boost utility even if efficiency is low. This is similar to the efficiency paradox discussed by Berg et al. (2019), who focus on the growth effect. This result can also encourage policy makers to prioritize projects with the highest complementarity with private factors, especially in countries where public investment inefficiency is a sizeable concern.

Given the crucial interconnection between investment efficiency and the complementarity between public capital and private sectors of production discussed in this chapter, empirical works linking estimates of this complementarity, public investment efficiency and economic outputs are needed. Beyond estimates discussed here, which shed some light on dominant features that emerge from worldwide analysis, more granular analysis could better inform on appropriate investment levels in low-efficiency countries. Complementarity can vary across sectors, and may depend on institutions. It may also depend on investment efficiency itself as low efficiency can be associated with a selection of projects which could have been done better by the private sector. In such a case, a critical remedy in low-efficiency countries would be to focus public investment on projects for which the case of complementarity with private factors of production is the strongest.

More broadly, public investment efficiency should remain a key area of research, as discussed in Kangur and Papageorgiou (2017). Key empirical and theoretical questions remain open. Effects of public investment efficiency changes on transitional growth could be explored further. Also, Kangur and Papageorgiou (2017) note that public investment efficiency could affect the level of technology, creating externalities and hence affecting growth even at the steady state in an endogenous growth model, as in Romer (1986) and Lucas (1988). This could reflect an effect of public R&D – which is part of public investment since the introduction of the 2008 System of National Accounts. Compared to physical investment, one possible difference is that basic research – the part of public R&D that is likely to less substitute for private research – may be non-rival, so that countries with inefficient public R&D may find it optimal to free ride and may hence prefer a low investment-to-GDP ratio even if complementarity is high. Further, the effect of public investment efficiency may interact with other factors. For instance, Buffie et al. (2012) show that low public investment efficiency can affect public debt sustainability. Finally, there is a lack of empirical work measuring variations in public investment efficiency over time.

NOTES

1. Jean-Marc Fournier and Fabien Gonguet were economists in the Fiscal Affairs Department of the IMF when writing this chapter. The authors thank Tewodaj Mogues, John Spray and Eivind Tandberg for useful comments and suggestions. The views expressed in this chapter are those of the authors and do not necessarily represent the views of the IMF, its Executive Board or IMF management.
2. Calderón et al. (2015) address this issue with estimates of a production function with a synthetic measure of infrastructure outputs instead of a monetary measure.
3. If public investment efficiency is improving, the public capital stock target should be higher and past stock affected by lower efficiency, quantitative output targets like kilometers of roads should thus be increased. Whether this should translate into an increase in public investment amounts or not is ambiguous as less public investment is required to achieve the same output objective.
4. With a similar Cobb-Douglas production function, Hulten (1996) also noted that efficiency affects the level rather than the growth rate.
5. Allais (1947) was largely unnoticed and the golden rule of accumulation of capital was popularized later by Phelps (1961, 1965). The presentation used here follows closely Phelps (1965).
6. In practice, the second condition would not hold only when efficiency of public capital is very close to zero.
7. In a steady-state growth path, private investment would depend on the provision of public infrastructure, so that a model with private investment could be nested in the model presented here – the function f introduced in equation (15.4) would embed the effect of public investment on output via private investment.
8. Berg et al. (2019) discuss threshold values for elasticities of substitutions with a nested CES production function including both public and private capital.
9. By contrast, Berg et al. (2019) assume that public investment to GDP ratio does not vary with efficiency. As it shall not vary with public investment efficiency in the Cobb-Douglas case anyway, in this case the results are similar.
10. An et al. (2019) cannot exclude that within advanced economies, elasticity of substitution between public and private capital could be one. However, this result is surrounded by a large uncertainty.

REFERENCES

Abiad, A., D. Furceri, and P. Topalova. 2016. ‘The Macroeconomic Effects of Public Investment: Evidence from Advanced Economies.’ *Journal of Macroeconomics* 50, pp. 224–40.

Agénor, P.-R. 2010. ‘A Theory of Infrastructure-Led Development.’ *Journal of Economic Dynamics and Control* 34, pp. 932–950.

Allain-Dupré, D. 2011. Multi-level Governance of Public Investment: Lessons from the Crisis. OECD Regional Development Working Papers No. 2011/05.

Allais, M. 1947. Economie et intérêt. (two volumes). Paris: Imprimerie Nationale.

Allen, R., M. Betley, C. Renteria and A. Singh. 2020. ‘Integrating Infrastructure Planning and Budgeting.’ In *Well Spent: How Strong Infrastructure Governance Can End Waste in Public Investment*, edited by G. Schwartz, M. Fouad, T. Hansen, and G. Verdier. Washington, DC: International Monetary Fund.

American Society of Civil Engineers (ASCE). 2018. ‘Infrastructure Report Card.’ ASCE, Reston, VA.

An, Z., A. Kangur and C. Papageorgiou. 2019. ‘On the Substitution of Private and Public Capital in Production.’ IMF Working Papers No. 2019/232, International Monetary Fund.

Arezki, R., P. Bolton, S. Peters, F. Samama, and J. Stiglitz. 2017. 'From Global Savings Glut to Financing Infrastructure.' *Economic Policy* 32 (90), pp. 221–261.

Arrow, K. J. and M. Kurz. 1970. *Public Investment, The Rate of Return, and-Optimal Fiscal Policy*. The Johns Hopkins University Press for Resources for the Future.

Aschauer, D. A. 1989. 'Is Public Expenditure Productive?' *Journal of Monetary Economics* 23, pp. 177–200.

Baum, A., T. Mogues, and G. Verdier. 2020. 'Getting the Most from Public Investment.' In *Well Spent: How Strong Infrastructure Governance Can End Waste in Public Investment*, edited by G. Schwartz, M. Fouad, T. Hansen, and G. Verdier. Washington, DC: International Monetary Fund.

Berg, A., E. F. Buffie, C. Pattillo, R. Portillo, A. F. Presbitero and L.-F. Zanna. 2019. 'Some Misconceptions About Public Investment Efficiency and Growth', *Economica* 86, pp. 409–430.

Berg, A., R. Portillo, S.-C. S Yang and L.-F. Zanna. 2013. 'Public Investment in Resource-Abundant Developing Countries.' *IMF Economic Review* 61, pp. 92–129.

Blazey, A., F. Gonguet, and P. Stokoe. 2020. 'Maintaining and Managing Public Infrastructure Assets.' In *Well Spent: How Strong Infrastructure Governance Can End Waste in Public Investment*, edited by G. Schwartz, M. Fouad, T. Hansen, and G. Verdier. Washington, DC: International Monetary Fund.

Bom P. R. D. and J. E. Ligthart. 2014. 'What Have We Learned From Three Decades Of Research On The Productivity Of Public Capital?' *Journal of Economic Surveys* 28 (5).

Braconier, H., and M. Pisu. 2013. Road Connectivity and the Border Effect: Evidence from Europe.

Buffie, E., A. Berg, C. Pattillo, R. Portillo, and F. Zanna. 2012. 'Public Investment, Growth and Debt Sustainability: Putting Together the Pieces.' *IMF Working Paper No. 12/144*. International Monetary Fund, Washington, DC.

Byrnes, P., Färe, R. and Grosskopf, S. 1984. 'Measuring Productive Efficiency: An Application to Illinois Strip Mines.' *Management Sciences* 30, pp. 671–681.

Calderón, C., E. Moral-Benito, and L. Servén. 2015. 'Is Infrastructure Capital Productive? A Dynamic Heterogeneous Approach.' *Journal of Applied Econometrics* 30, pp. 177–198.

Canning, D. 1998. 'A Database of World Stocks of Infrastructure.' *The World Bank Economic Review* 12 (3), pp. 529–547.

Canning, D. and E. Bennathan. 2000. 'The Social Rate of Return on Infrastructure Investments.' Policy Research Working Paper Series 2390. The World Bank.

Chaponda, T, C. Matsumoto, and L. K. Murara. 2020. 'The Public Investment Management Assessment Framework: An Overview.' In *Well Spent: How Strong Infrastructure Governance Can End Waste in Public Investment*, edited by G. Schwartz, M. Fouad, T. Hansen, and G. Verdier. Washington, DC: International Monetary Fund.

Charnes, A., W. W. Cooper, and E. Rhodes. 1978. 'Measuring the Efficiency of Decision-Making Units.' *European Journal of Operational Research* 2, pp. 429–444.

Chirinko, R. S., and D. Mallick. 2017. 'The Substitution Elasticity, Factor Shares, and the Low-Frequency Panel Model.' *American Economic Journal: Macroeconomics* 9 (4), pp. 225–253

Collier, P. 2007. *The Bottom Billion: Why the Poorest Countries are Failing and What Can Be Done About It.* Oxford University Press, New York.

Collier, P., M. Kirchberger, and M. Söderbom. 2016. 'The Cost of Road Infrastructure in Low- and Middle-Income Countries.' *World Bank Economic Review* 30 (3) pp. 522–548.

Dabla-Norris, E., J. Brumby, A. Kyobe, Z. Mills, and C. Papageorgiou. 2012. 'Investing in Public Investment: An Index of Public Investment Efficiency.' *Journal of Economic Growth* 17 (3), pp. 235-266.

De Long, J. B., and L. H. Summers. 1992. 'Equipment Investment and Economic Growth: How Strong is the Nexus?' *Brookings Papers on Economic Activity* 1992 (2), pp. 157–211.

Donaubauer, J., B. E. Meyer and P. Nunnenkamp. 2016. 'A New Global Index of Infrastructure: Construction, Rankings and Applications.' *The World Economy* 39 (2), pp. 236–259.

Dotti, N. F. 2015. 'Unwritten Factors Affecting Structural Funds: The Influence of Regional Political Behaviours on the Implementation of EU Cohesion Policy.' *European Planning Studies* 24 (3), pp. 530–550.

Fernald, J. G. 1999. 'Roads to Prosperity? Assessing the Link between Public Capital and Productivity.' *American Economic Review* 89 (3), pp. 619–638.

Flyvbjerg, B. 2009. 'Survival of the Unfittest: Why the Worst Infrastructure Gets Built—And What We Can Do about It.' *Oxford Review of Economic Policy* 25 (3), pp. 344–367.

Fournier, J.-M. 2016. 'The Positive Effect of Public Investment on Potential Growth.' *OECD Economics Department Working Paper No. 1347*.

Furceri, D., and B. G. Li. 2017. 'The Macroeconomic (and Distributional) Effects of Public Investment in Developing Economies.' *IMF Working Paper No. 17/217*. International Monetary Fund, Washington, DC.

Gechert, S., T. Havránek, Z. Havránková, and D. Kolcunova. 2019. 'Death to the Cobb-Douglas Production Function.' *IMK Working Paper No. 201*. Hans-Böckler-Stiftung, Institut für Makroökonomie und Konjunkturforschung (IMK), Düsseldorf.

Gupta, S., A. Kangur, C. Papageorgiou, and A. Wane. 2014. 'Efficiency-Adjusted Public Capital and Growth.' *World Development* 57 (C), pp. 164–178.

Gurara, D., K. Kpodar, A. F. Presbitero, and D. Tessema. 2020. 'On the Capacity to Absorb Public Investment: How Much Is Too Much?' *IMF Working Paper No. 20/48*. International Monetary Fund, Washington, DC.

Hulten, C. 1996. 'Infrastructure Effectiveness as a Determinant of Economic Growth: How Well You Use it May Be More Important than How Much You Have.' *NBER Working Paper No. 5847* (revised December 2005).
Hurlin, C., and F. Arestoff. 2010. 'Are Public Investments Efficient in Creating Capital Stocks in Developing Countries?' *Economics Bulletin* 30 (4), pp. 3177–3187.
International Monetary Fund (IMF). 2015. *Making Public Investment More Efficient*. Washington, DC: IMF.
———. 2017. *Investment and Capital Stock Dataset*.
———. 2018. 'Public Investment Management Assessment: Review and Update.' IMF Policy Paper, Washington, DC.
———. 2019. *Curbing Corruption*. Fiscal Monitor, April.
———. 2020. *Public Investment for the Recovery*. Fiscal Monitor, October.
Isham, J., and D. Kaufmann. 1999. 'The Forgotten Rationale for Policy Reform: The Productivity of Investment Projects.' *Quarterly Journal of Economics* 114 (1), 149–184.
Kamps, C. 2006. 'New Estimates of Government Net Capital Stocks for 22 OECD Countries, 1960-2001.' *IMF Staff Papers* 53 (1), pp. 120–150.
Kangur, A. and C. Papageorgiou. 2017. 'Public Sector Investment Efficiency in Developing Countries.' *The New Palgrave Dictionary of Economics*, pp. 1–10. London, UK: Palgrave Macmillan.
Kim, J.-H., J. A. Fallov, and S. Groom. 2020. *Public Investment Management Reference Guide. International Development in Practice*. Washington, DC: World Bank.
King, R. G., and R. Levine. 1994. 'Capital Fundamentalism, Economic Development, and Economic Growth.' *Carnegie-Rochester Conference on Public Policy* 40, pp. 259–292.
Kumbhakar, S. C., C. F. Parmeter and V. Zelenyuk. 2020. 'Stochastic Frontier Analysis: Foundations and Advances I.' In S. C. Ray et al. (eds.), *Handbook of Production Economics*. Springer Nature Singapore Pte Ltd.
Lucas, R. E., Jr. 1988. 'On the Mechanics of Economic Development.' *Journal of Monetary Economics*, 22 (4), pp. 3–42.
Marglin S. A. 1963. 'The Opportunity Costs of Public Investment.' *The Quarterly Journal of Economics* 77 (2), pp. 274–289.
Mauro, P.. 1998. 'Corruption and the Composition of Government Expenditure.' *Journal of Public Economics* 69, pp. 263−279.
Melina, G., S.-C. S.Yang and L.-F. Zanna. 2016. 'Debt Sustainability, Public Investment, and Natural Resources in Developing Countries: The DIGNAR Model.' *Economic Modelling* 52 (B), pp. 630−649.
Miyamoto, H., A. Baum, N. Gueorguiev, J. Honda, and S. Walker. 2020. 'Growth Impact of Public Investment and the Role of Infrastructure Governance.' In *Well Spent: How Strong Infrastructure Governance Can End Waste in Public Investment*, edited by G. Schwartz, M. Fouad, T. Hansen, and G. Verdier. Washington, DC: International Monetary Fund.
Murillo-Zamorano, L. R. 2004. 'Economic Efficiency and Frontier Techniques.' *Journal of Economic Surveys* 18 (1) pp. 33–76.
OECD. 2017. *Getting Infrastructure Right: A Framework for Better Governance*. Paris: OECD Publishing.
Pattanayak, S., and C. Verdugo-Yepes. 2020. 'Protecting Public Infrastructure from Vulnerabilities to Corruption: A Risk-Based Approach.' In *Well Spent: How Strong Infrastructure Governance Can End Waste in Public Investment*, edited by G. Schwartz, M. Fouad, T. Hansen, and G. Verdier. Washington, DC: International Monetary Fund.
Phelps, E. S. 1961. 'The Golden Rule of Accumulation.' *The American Economic Review* 51, pp. 638−643.
Phelps, E. S. 1965. 'Second Essay on the Golden Rule of Accumulation.' *The American Economic Review* 55 (4), pp. 793–814.
Presbitero, A. 2016. 'Too Much and Too Fast? Public Investment Scaling-Up and Absorptive Capacity.' *Journal of Development Economics* 120, pp. 17–31.
Pritchett, L. 2000. 'The Tyranny of Concepts: CUDIE (Cumulated, Depreciated, Investment Effort) Is Not Capital.' *Journal of Economic Growth,* 5 (4), pp. 361–384.
Rajaram, A., T. M. Le, K. Kaiser, J.-H. Kim, and J. Frank (editors). 2014. *The Power of Public Investment Management: Transforming Resources into Assets for Growth*. Washington, DC: World Bank.
Romer, P. M. 1986. 'Increasing Returns and Long Run Growth.' *Journal of Political Economy*, 94 (5), pp. 1002–1037.
Schwartz, G., M. Fouad, T. Hansen, and G. Verdier, editors. 2020. *Well Spent: How Strong Infrastructure Governance Can End Waste in Public Investment.* Washington, DC: International Monetary Fund.
van der Ploeg, F. 2012. 'Bottlenecks in Ramping up Public Investment.' International Tax and Public Finance, Springer. *International Institute of Public Finance*, 19 (4), pp. 509–538.
World Bank. 2014. *World Development Report 2014: Risk and Opportunity Managing Risk for Development*. Washington, DC: World Bank.

PART IV

PUBLIC SECTOR EFFICIENCY ACROSS THE WORLD

16. Efficiency and economic growth: A panel analysis of Colombian regions

Luis Alberto Gutiérrez-Arango, Víctor Giménez, Daniel Osorio-Barreto, and Diego Prior

1. INTRODUCTION

This chapter analyzes the factors that generate economic growth and social progress in Colombia, a good example of a developing country. Two perspectives are taken into consideration: i) the potential convergence in rates of growth in the different regions,[1] and ii) the dependence of growth on the availability of public infrastructures.

Regarding the first of these perspectives, sustainable growth rates need to be convergent in order to reduce economic differences among regions and thereby avoid political instability and armed conflicts, a situation that has defined Colombia's recent history (González et al., 2022).

During the late 1980s, economic convergence was observed with the slowdown in growth in countries such as the United States, while others, such as Japan, grew at high rates. Economic convergence, understood as the process by which countries with less capital per employee tend to grow faster in per capita terms than those with a higher capital/labor ratio (Sala-i-Martín, 2000), is the result of the traditional approach to measuring economic growth developed by Solow (1957). A similar research agenda focused on productivity and postulated that the relative slowdown in productivity in the United States and European countries may be due to a natural process of convergence; in this way, countries with a lower level of productivity can catch up to those with a high level of productivity (Wu, 2000).

Baumol (1986) considered the empirical results for developing countries and found an apparent trend toward convergence within some groups of countries (industrialized, intermediate and less developed) but not between these groups. In fact, convergence is only observed in the group of industrialized economies, but there is no clear pattern for the rest of the groups. Considering the case of less developed countries, Baumol (1986) observed that the transmission process for knowledge, which helps the catching-up process, does not appear to be present, which causes divergence in the rates of economic growth. A similar conclusion was obtained in a more recent study (Juknys et al., 2017), suggesting that inequity is persistent and growing in developing economies. Considering these results, our objective here is to verify the extent to which divergence in economic growth is present among Colombian regions with different levels of development. To answer this question, we use frontier methods to estimate the basic components of the total factor productivity change (technological change and technical efficiency change) that can have a direct impact on the levels of economic growth (see foundations about efficiency in Farrell (1957)).

From the second perspective, as economic growth is dependent on the consumption of inputs, the efficient provision of public capital to provide infrastructure services is one of the most important tools for development policies, especially in emerging economies. The absence of adequate infrastructures, as well as the inefficient provision of infrastructure services, therefore, constitute major obstacles to the effective implementation of development policies to obtain successful economic growth rates. Countries need to expand and modernize their basic infrastructure to achieve maximum levels of coverage across all their territories. This will help to provide the infrastructures that economic agents and individuals require. Here lies our second orientation: private and public capital investments are considered as driving factors for economic growth in the estimations.

The technique used is an evolution of Nishimizu and Page (1982) and Wu (2000), which allows the decomposition of productivity growth into technical efficiency change and technological progress. Our proposal defines a stochastic frontier econometric model, applied to 23 Colombian regions for the years 1996 to 2009. Our results contribute a new perspective to the literature, as the joint consideration of private and public capital investment is not common in developing countries.

Results show that Colombian regions do not converge in terms of output growth caused by technical efficiency change. These results are significant before and after considering the availability of public capital. However, the gap is not growing because of the incapacity for big regions to generate technological progress. The obvious implication is that plans must be put in place to bring down levels of inefficiency in less advanced regions and to promote more consistent technological progress.

2. THEORETICAL FRAMEWORK

To estimate the levels of technical efficiency, this research defines a stochastic frontier translog production function. Unlike the traditional growth accounting method, this orientation makes it possible to lower the potential output level due to the presence of inefficiency (Wu, 2000). The panel data version for this model can be expressed as follows:

$$y_{it}^{f} = f(x_{it}, t), \quad t = 1, \ldots, T; \; i = 1, \ldots, N. \tag{16.1}$$

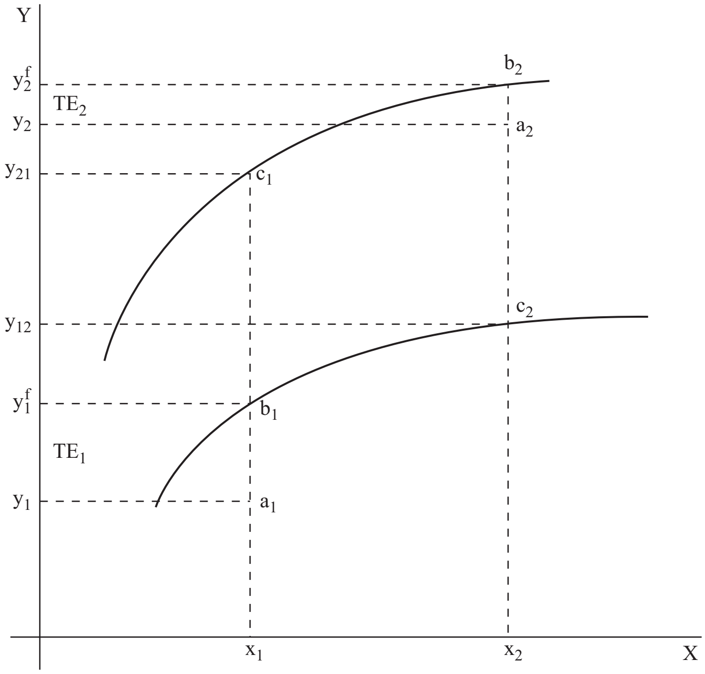
The first term of the equation refers to the level of production at the frontier for region i in year t, based on the existing technological knowledge. Now, the observed product or y_{it}, considering the available levels of inputs x_{it} and the potential inefficiency, can be represented mathematically as follows:

$$Y_{it} = y^{f} .TE_{it} = f(x_{it}, t) .TE_{it}. \tag{16.2}$$

where TE_{it} is the technical efficiency coefficient, which can be defined as the quotient of the observed output over the potential level of output at the efficient frontier. Using derivatives with respect to time, the above equation can be transformed into:

$$\dot{y}_{it} = f_x \dot{x}_{it} + f_t + \dot{TE}_{it} \tag{16.3}$$

where the dotted variables stand for the derivatives with respect to time; f_t and f_x are the elasticities of the product with respect to t and a x.



Source: Authors after Wu (2000).

Figure 16.1 Decomposing the driving factors of output growth

With regard to equation (16.3), and in relation to the traditional growth accounting method, Solow (1957) attributed the growth in production to the increase in the level of input consumption and to technological change. However, equation (16.3) enriches the Solow dichotomy by attributing the progress of the observed product to a movement along a path at or below the production frontier (input growth), a movement to or from the production frontier (change in technical efficiency) and changes in the production frontier (technological progress or regress). This process is described graphically in Figure 16.1.

Points a_1 and a_2 in Figure 16.1 represent the observed level of the production of y_1 and y_2, respectively, for a specific region, corresponding to the time periods 1 and 2. Additionally, best practice outputs are represented by the points b_1 and b_2. The difference between the observed output and its respective best practice is the technical inefficiency. Changes between the two time periods, measured as the difference between TE_1 and TE_2, provide the change in technical efficiency. Mathematically, it can be represented as follows:

$$\begin{aligned}\Delta y = (y_2 - y_1) &= \left(y_2^f - TE_2\right) - \left(y_1^f - TE_1\right) = \left(y_2^f - y_1^f\right) + \left(TE_1 - TE_2\right) \\ &= \left(y_2^f - y_{12}\right) + \left(y_{12} - y_1^f\right) + \left(TE_1 - TE_2\right).\end{aligned} \tag{16.4}$$

According to identity (16.4), the variation in the output level ($y_2 - y_1$) can be decomposed into: (1) output growth generated by the technological change (progress or regress)

$(\dot{TCH} = y_2^f - y_{12})$, (2) output growth derived from the increase in the input consumption $(y_{12} - y_1^f)$ and (3) output growth due to the change in technical efficiency $\dot{TE} = (TE_1 - TE_2)$. Equation (16.4) indicates that the variation in the total factor productivity, which is understood as the growth of the product that is not being explained by the variation in inputs, can be expressed as follows:

$$\dot{TFP}_{it} = \dot{TCH}_{it} + \dot{TE}_{it}, \ i = 1, \ldots, N; \quad t = 1, \ \ldots, T. \tag{16.5}$$

Equation (16.5) is the cornerstone of our empirical work, since conventional techniques of accounting for economic growth, such as that of Solow (1957), do not distinguish between technological change and changes in technical efficiency. Following Nishimizu and Page (1982), these components are analytically distinct and can have quite different implications from the policy making perspective. Faced with technological progress, it can be interpreted as the result of the innovation of the best practice frontier, while technical efficiency change, among other possible causes, can be expressive of the convergence process (Wu, 2000).

3. LITERATURE REVIEW

This section presents the results of a systematic literature review. The objective is to identify, after the seminal work of Nishimizu and Page (1982), the most relevant empirical applications that have contributed to the research field using the productive efficiency approach. Based on the above, five key connectors are identified to include in the search engines, which are applied in the Web of Science (WoS) and Scopus platforms. The keywords used are: *Productivity* and *Technical Efficiency*, *Technological Progress*, *Frontier* and *Economic Growth*.

The search equation with these keywords identified 32 articles (published in 30 academic journals between 2000 and 2020) that include our keywords in their title or abstract. These articles each received an average of 13.59 citations. We then identified which estimation method, parametric or non-parametric, the authors of these papers used to determine the drivers of economic growth. Table A16.1 in the Appendix contains this information. It appears that stochastic parametric methods are the most common estimation method when accounting for economic growth.

Given our empirical orientation, we proceeded to analyze in detail the scientific articles that have used the parametric approach. Wu (2000) analyzes the overall Chinese economy, and was one of the first scholars to take this orientation, since previous studies focused on the agricultural and industrial sectors. The author uses a stochastic frontier model with a translog production function to estimate the behavior of productivity, and decomposes it into technological progress and changes in technical efficiency, for Chinese provinces, using regional GDP as an output variable for the years 1981 and 1995. The study concludes that the results of technical efficiency change for the provinces of China have converged rapidly since the early 1980s as a result of the economic reform that took place at the end of the 1970s. Özyurt and Guironnet (2011) draw a similar conclusion, arguing that, during the period 1994–2006, the results indicate a surprising trend toward convergence among the Chinese provinces, which tend to compensate the negative scale effects by increasing production efficiency through technological progress.

Similarly, Li and Liu (2011) examine economic growth and total factor productivity in the Chinese provinces after the reform carried out in the late 1970s. To do this, they use a stochastic frontier model with a translog production function incorporating human capital, measured by the years of schooling of the population. The main objective of the study is to estimate the attributes of economic growth for the years between 1986 and 2006 which, according to the authors, are: i) input growth, ii) the adjusted scale effect, iii) technical progress and iv) efficiency growth. The results of the study indicate that the growth in inputs is the main driver of economic growth, and the availability of human capital is inadequate, although it has a positive effect on the output growth. Additionally, the technological progress explains an important percentage of productivity growth.

Referring to the role of human capital in the Chinese provinces, Zhou et al. (2011) find, for the period 1985–2008, that the average elasticity of labor production is greater than the other two inputs of capital and human capital. In addition, the effects over time of technical efficiency in the post-reform Chinese economy is significantly dependent on the growth of inputs, highlighting differences in efficiency levels across regions.

In their study of 15 former Soviet Union countries, Arazmuradov et al. (2014) use a stochastic frontier for the period 1995–2008. They find that human capital improves technical efficiency. In addition, human capital has a positive impact on total factor productivity, which in turn stimulates growth in the level of GDP.

A more recent study by Brock and Ogloblin (2018), using a stochastic frontier model and a translog production function for the Russian economy, examines technical inefficiency change, technological progress and returns to scale during the cyclical expansion of 1998–2007. Their objective is to include both branches of activity that contain the market and the non-market economies. The services sector, together with the production of goods with a high intensity of skilled labor, has proven to be relatively more efficient than traditional manufacturing Soviet-era production sectors. In the time window analyzed, technical efficiency change decreased notably during expansion, while technological progress increased (Brock and Ogloblin, 2018).

Han et al. (2002), study the sources of economic growth between 1987 and 1993 for Hong Kong, Singapore, Japan and South Korea. They estimate a Cobb-Douglas stochastic frontier production function using information from 20 manufacturing sectors. The approach used allowed the authors to decompose the growth of total factor productivity into changes in technical efficiency and technological progress. The results identify the positive relationship between the increase in inputs as the main source of economic growth in the countries of East Asia. In addition, they provide statistical evidence to support positive changes in technical efficiency. For a specific region, the northeast of China, Ganzhou (2011) estimates a panel of data for the period 1978–2008, concluding that the contribution of input growth gradually reduces its impact on the level of output growth, while the contribution of total factor productivity becomes stronger. Similarly, technological progress increases continuously, and technical efficiency grows slowly, a scenario that can impact the sustainable growth of the region.

Liao et al. (2007) study the behavior of the manufacturing sector using a stochastic frontier model and specifying a translog production function for the economies of South Korea, the Philippines, Hong Kong, Indonesia, Malaysia, Singapore and Taiwan between 1963 and 1998. Their article focuses on analyzing technological progress, the behavior of technical efficiency and the change in productivity in economic growth. From the results, the authors conclude that, although the accumulation of higher inputs is the main source of output growth, the positive variation in total factor productivity represents an increasing proportion of the output growth, explained by the advances made in technical efficiency change. Shao and Lin (2016) assess the production performance for information technology services in twelve OECD countries in the years 2000 to 2011, finding an annual growth rate of 7.4 percent in production in this activity. The behavior of the information services industry varies in each country, resulting in strengths that become sources of competitive advantage, or weaknesses that could be drivers to improve performance.

In their Chinese case study, Shi et al. (2017) estimate the technical efficiency of the forestry sector and the growth of total factor productivity for the period 2004–15, using a stochastic frontier model. They find general inefficiency in forestry production and an average technical efficiency of 0.546 during the time period considered. Similarly, Liu and Tsai (2021) define a stochastic frontier analysis approach and a flexible translog production function, considering a neutral technological progress. The research aimed to evaluate the behavior of technical efficiency, technological and scale change, as well as the change in total factor productivity. The study explores the convergence in the Chinese hotel industry with star ratings in 31 provinces, municipalities and regions for the years 2001–15. The results indicate that the change in total factor productivity in the hotel industry was generally favorable and was driven by both technical efficiency change and technological change.

Another study that analyzes the changes in productivity between countries, and that breaks this down into technical efficiency and technological progress, is that of Lee and Cheng (2011). This article studies the cases of Philippines, Singapore, Malaysia, Thailand and Indonesia, using a stochastic frontier model with the specific temporal pattern for technical efficiency for the period 1981–2003. The authors define an estimation of a unique temporal pattern of changes in productivity in each country. The empirical results indicate that, during the study period, growth in Singapore and Malaysia was largely driven by technological progress and the accumulation of inputs, while growth in Thailand derived from an improvement in the technical efficiency change and by the growth of input consumption.

Colino et al. (2014) analyze the determinants of the increase in total factor productivity for 26 OECD member countries between 1965 and 2010, studying technical efficiency and technological progress. Taking as a reference the economies of the European Union, they found different patterns of productivity growth between world technology leaders and countries with low initial levels of productivity, implying that changes in efficiency may be the main result of the evolution of the stock of knowledge in technologically dependent advanced economies.

For the European Union (EU), Brock and Ogloblin (2014) use a stochastic frontier model to econometrically estimate how technical efficiency, technological progress and returns to scale contributed to the economic growth of EU member countries for the years 1979–2000, including years of schooling as a proxy for human capital. They find that technical inefficiency, in general, is low. However, it diverges over time, with almost no displacement of the aggregate frontier.

For a 30-year time frame (1980–2010), and a sample of 140 countries,[2] as well as incorporating both physical and human capital and labor into the production function, Mamonov and Pestova (2015) add the aggregate index of institutional development designed by the Fraser Institute and the *WDI* indicators of infrastructure conditions. The results show faster technological progress in developing countries than in the more developed economies.

In summary, we found a gap in the literature for articles estimating total factor productivity using the stochastic frontier approach for Latin American countries. In this context, our application to the case of Colombia is particularly significant because of the economic and political events of recent decades.

4. MODEL TO ESTIMATE

To meet the objectives set out in the introduction, we will define a translog production function that can be estimated using a stochastic frontier process. This production function is supposed to observe a non-neutral technological change (as the change in the marginal rates of input substitution depends on the existing levels of *L* and *K*). The precise definition of the translog production function[3] is as follows:

$$\ln Y_{it} = \alpha_0 + \alpha_1 \cdot t + \frac{1}{2} \cdot \alpha_2 \cdot t^2 + \left(\beta_0 + \beta_1 \cdot t\right) \cdot \ln L_{it} + \left(\gamma_0 + \gamma_1 \cdot t\right) \cdot \ln K_{it} + \frac{1}{2} \cdot \left(\eta_1 \cdot \ln L_{it}^2 + 2 \cdot \eta_2 \cdot \ln L_{it} \cdot \ln K_{it} + \eta_3 \cdot \ln K_{it}^2\right) + \varepsilon_{it} \tag{16.6}$$

under the following restrictions:

$$\varepsilon_{it} = v_{it} + u_{it} \tag{16.7}$$

where, *Y*, *L* and *K* stand for output, labor and capital inputs, respectively. In addition, α_i, β_j, γ_k and η_l are parameters to be estimated. The error term ε_{it} combines white noise, v_{it}, and the term associated with technical inefficiency, u_{it}. It is assumed that v_{it} and u_{it} are independently distributed, and have normal distributions with zero means and constant variations, σ_u^2 and σ_v^2.

As indicated in the introduction, the originality of our proposal is the introduction of public capital, so the original translog production function is transformed as follows:

$$\begin{aligned}\ln Y_{it} = {} & \alpha_0 + \alpha_1 \cdot t + \frac{1}{2} \cdot \alpha_2 \cdot t^2 + \left(\beta_0 + \beta_1 \cdot t\right) \cdot \ln L_{it} + \left(\gamma_0 + \gamma_1 \cdot t\right) \cdot \ln K_{it} + \\ & + \left(\tau_0 + \tau_1 \cdot t\right) \cdot \ln G_{it} + \frac{1}{2} \cdot \left(\eta_1 \cdot \ln L_{it}^2 + \eta_2 \cdot \ln L_{it} \cdot \ln K_{it} + \eta_3 \cdot \ln K_{it}^2 + \eta_4 \cdot \ln G_{it}^2 + \right. \\ & \left. + \eta_5 \cdot \ln L_{it} \cdot \ln G_{it} + \eta_6 \cdot \ln K_{it} \cdot \ln G_{it}\right) + \varepsilon_{it}\end{aligned} \tag{16.8}$$

where G_{it} stands for the stock of public capital for unit *i* in time period *t*.

Equations (16.6) and (16.8) are estimated by maximum likelihood. In these models, following Battese and Coelli (1992), the inefficiency component is modeled as time variant:

$$u_{it} = exp\{-h(T - t_i)\}\ u_i \tag{16.9}$$

where $u_i \sim N^+\left(u, \sigma_u^2\right)$

In equation (16.9), when $\eta > 0$, the degree of efficiency increases over time and when $\eta < 0$ the degree of efficiency decreases over time. In the last period $t_i = T$, then, inefficiency is in base-level $(u_{iT} = u_i)$.

Equation (16.10) indicates the variation in technical efficiency:

$$\dot{TE}_{it} = u_{it-1} - u_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ \ldots, T. \tag{16.10}$$

Likewise, technological change, for the translog production function of equation (16.8), can be obtained from the derivative of the output with regard to time:

$$T\dot{C}H_{it} = \alpha_1 + \alpha_2 \cdot t + \beta_1 \cdot \ln L_{it} + \gamma_1 \cdot \ln K_{it} + \tau_1 \cdot \ln G_{it}. \qquad (16.11)$$

5. DATA

To estimate the models proposed, *GDP* is taken as output while total number of employees (*L*), private capital (*K*) and public capital stock (*G*) are the variables taken as inputs for the Colombian regions. *GDP*, *K* and *G* are taken from Vallecilla (2011), following the guidelines of the Simplified System of Regional Accounts (*SSCD*) in version 3 of the National Accounts System, updated to 2008. The time frame spans 14 years, beginning in 1996 and ending in 2009. The number of employees was taken from National Statistics Department (*DANE*) (2020) information on the annual average of employed persons per region obtained through the Large Integrated Household Survey (*GEIH*).

Table A16.2 in the Appendix reports descriptive statistics, at regional level, for these variables. There are observable differences between the three large regions (Bogotá, D.C., Antioquia and Valle) and the rest. These three regions account for around 55 percent of the total GDP.

Regarding capital investments, the descriptive statistics show considerable differences among regions. While in the large regions private capital represents around 35 percent of the total investments (54 percent in Antioquia), in small regions this percentage is substantially lower (around 9 percent in Chocó). This means that estimations that do not incorporate the presence of public capital investments—as in the original model represented in equation (16.6)—may be biased because they fail to account for a relevant input variable. In these circumstances, the extended model—presented in equation (16.8)—appears to be a better approximation of the Colombian production process.

6. RESULTS

The results of the estimations specified in equations (16.6) and (16.8) are presented in Table 16.1. Variables in the model were scaled to have unit means, so that the first-order coefficients of the translog functions can be interpreted as elasticities of output with respect to inputs evaluated at the sample means (Coelli et al., 2005); this also serves to reduce multicollinearity in polynomial or interaction-effect models (Hamilton, 2012).

The models appear to be plausible and quite stable without significant changes in signs. The sum of first-order coefficients (β_0, γ_0 and τ_0 in the extended model) is less than one in both models, so decreasing returns to scale are present at the sample mean. It is also confirmed with statistical significance that technical progress seems to decrease with the magnitude of private capital input (as $\gamma_1 < 0$) but not with the level of public capital input (because, without having statistical significance, τ_1 is very close to zero). Regarding the characteristics of the inputs, labor and private capital appear to be complementary inputs (as observed in the sign and significance of η_2, but only in the original model). Additionally, coefficients related with public capital (τ_1 and η_4) indicate that the elasticity of this input is very low.[4] These results are in line with Holtz-Eakin (1994), Pereira and De Frutos (1999) or Montolio and Solé-Ollé (2009), expressing a potential negative effect of congestion. Finally, to compare the two models, we conduct the likelihood-ratio test finding that the extended model fits better with the data than the original model.[5]

The discrete variables in the model were scaled to have unit means so that the first-order coefficients of the translog function can be interpreted as elasticities of output with respect to inputs evaluated at the sample means (Coelli et al. 2005).

The estimations allow us to determine the specific level of technical efficiency change $(\dot{TE}_{it})$, the technological change $(T\dot{C}H_{it})$, which may be technological progress or regress, and the impact of total factor productivity on the output growth.

The average values corresponding to $\dot{TE}_{it}$ for the 23 Colombian regions (according to the original model and extended models) are presented in Table 16.2. The general picture is that technical efficiency decreases over time (as η is negative and significant in both models). It appears that regional disparities are sustained as interior regions with larger size present better results than those located on the coast. This means that smaller regions find it more difficult to reach their reference frontier, which hinders convergence in terms of output growth. The list of fastest growing regions based on technical efficiency change is very stable, which indicates that the potential for further improvements may be exhausted, so future economic growth should be based on innovations.

Table 16.1 Results of the estimation translog production function

	Original model (16.6) coefficients	Extended model (16.8) coefficients
α_0	0.263***	0.313***
α_1	0.013***	0.014**
α_2	0.006***	0.006***
β_0	0.426***	0.337**
β_1	−0.006	−0.007
γ_0	0.240***	0.318***
γ_1	−0.007***	−0.006**
τ_0	-	0.082
τ_1	-	0.004
η_1	0.011	0.278
η_2	0.146*	−0.013
η_3	0.010	0.062*
η_4	-	−0.438***
η_5	-	0.149
η_6	-	0.014

Notes: (*) $p < 0.1$; (**) $p < 0.05$; (***) $p < 0.001$.

Source: Authors' calculations.

Table 16.2 Technical efficient change

	Original model (16.6)			Extended model (16.8)		
Region	Mean	Min	Max	Mean	Min	Max
Santander	0.000	0.000	0.000	−0.001	−0.001	0.000
Valle	−0.001	−0.001	0.000	−0.004	−0.005	−0.001
Antioquia	−0.001	−0.001	−0.001	−0.003	−0.003	−0.001
Cundinamarca	−0.001	−0.001	−0.001	−0.004	−0.005	−0.001
Bogotá, D.C.	−0.002	−0.002	−0.002	−0.001	−0.002	−0.002
Bolivar	−0.002	−0.002	−0.002	−0.004	−0.005	−0.002
Atlántico	−0.003	−0.004	−0.003	−0.009	−0.011	−0.003
Boyacá	−0.004	−0.004	−0.004	−0.010	−0.012	−0.004
Tolima	−0.005	−0.005	−0.004	−0.008	−0.010	−0.005
Meta	−0.005	−0.005	−0.005	−0.010	−0.012	−0.005
Huila	−0.005	−0.005	−0.005	−0.011	−0.014	−0.005
Cesar	−0.006	−0.006	−0.005	−0.001	−0.001	−0.006
Magdalena	−0.006	−0.006	−0.006	−0.002	−0.003	−0.006
Norte de Santander	−0.006	−0.007	−0.006	−0.012	−0.015	−0.006
Nariño	−0.006	−0.007	−0.006	−0.010	−0.012	−0.006
Caldas	−0.007	−0.008	−0.007	−0.015	−0.019	−0.007
Risaralda	−0.007	−0.008	−0.007	−0.015	−0.019	−0.007
Córdoba	−0.008	−0.008	−0.007	−0.013	−0.016	−0.008
La Guajira	−0.008	−0.008	−0.007	−0.013	−0.016	−0.008
Cauca	−0.009	−0.010	−0.009	−0.022	−0.028	−0.009
Quindío	−0.012	−0.012	−0.011	−0.017	−0.021	−0.012
Sucre	−0.014	−0.015	−0.014	−0.017	−0.021	−0.014
Chocó	−0.022	−0.024	−0.021	−0.029	−0.036	−0.022

Source: Authors' calculations.

Results from the extended model for $\dot{TE}_{it}$ coincide. Hence, changes in the levels of technical efficiency do not help to produce convergence in the rates of output growth because smaller regions have problems catching up to their reference frontier (Abramovitz, 1986). These movements generate severe difficulties for them to converge in terms of output growth.

Summing up, results indicate that the lack of convergence is confirmed for both the original and the extended models.

We now turn to the levels of technological change ($T\dot{C}H_{it}$). Table 16.3 presents the results from the original and the extended models. Contrary to what is shown in Table 16.2, it appears that smaller regions are able to obtain greater technological change than larger ones. In fact, Valle, Antioquia and Bogotá, D.C. generate a negative, close to zero, average level of technological change. This situation is similar in both scenarios, the original and the extended

Table 16.3 Technological change

Region	Original model (16.6)			Extended model (16.8)		
	Mean	Min	Max	Mean	Min	Max
Chocó	0.061	0.025	0.101	0.048	0.012	0.088
La Guajira	0.057	0.021	0.092	0.053	0.018	0.087
Quindío	0.045	0.009	0.082	0.037	0.000	0.074
Sucre	0.043	0.006	0.082	0.033	−0.004	0.072
Cesar	0.041	0.004	0.077	0.033	−0.005	0.070
Huila	0.037	0.001	0.073	0.034	−0.001	0.071
Meta	0.036	0.002	0.070	0.034	−0.001	0.068
Nariño	0.035	0.000	0.071	0.028	−0.008	0.064
Norte de Santander	0.034	−0.002	0.071	0.029	−0.006	0.066
Magdalena	0.034	0.000	0.070	0.027	−0.008	0.063
Risaralda	0.029	−0.006	0.066	0.025	−0.010	0.062
Tolima	0.028	−0.008	0.065	0.024	−0.011	0.061
Cauca	0.027	−0.005	0.061	0.024	−0.008	0.057
Boyacá	0.027	−0.010	0.065	0.027	−0.010	0.065
Caldas	0.027	−0.010	0.065	0.023	−0.014	0.061
Córdoba	0.025	−0.009	0.061	0.020	−0.014	0.056
Bolivar	0.017	−0.018	0.053	0.015	−0.020	0.051
Santander	0.016	−0.020	0.053	0.018	−0.019	0.055
Cundinamarca	0.013	−0.022	0.048	0.013	−0.023	0.048
Atlántico	0.012	−0.022	0.047	0.010	−0.023	0.046
Valle	0.002	−0.033	0.038	0.004	−0.031	0.041
Antioquia	−0.003	−0.038	0.034	0.003	−0.032	0.040
Bogotá, D.C.	−0.012	−0.047	0.024	−0.004	−0.038	0.032

Source: Authors' calculations.

models. These results indicate that more advanced regions seem to have exhausted their tools and capacities to innovate in order to generate technological progress. Overall results are in line with those described by Özyurt and Guironnet (2011), where $\dot{TE}_{it}$ component decreases while $\dot{TCH}_{it}$ increases (see also Brock and Ogloblin, 2014, 2018).

Table 16.4 presents the aggregated results of the two components on the levels of output growth due to the Total Factor Productivity Change ($\dot{TFP}_{it}$, as formalized in equation 16.5). In both cases we see how the impact of technological change dominates the movements of technical efficiency change, so the classification of Colombian regions according to their movements on Total Factor Productivity does not differ substantially from what is presented in Table 16.3. These results indicate that more advanced regions seem to be unable to compensate for the lack technological progress by improving their levels of technical efficiency change.

Table 16.4 Total factor productivity change

	Original model (16.6)			Extended model (16.8)		
Region	Mean	Min	Max	Mean	Min	Max
La Guajira	0.049	0.013	0.085	0.04	0.002	0.079
Chocó	0.039	0.001	0.08	0.019	−0.024	0.066
Cesar	0.035	−0.002	0.072	0.032	−0.006	0.064
Quindío	0.033	−0.003	0.071	0.02	−0.021	0.062
Huila	0.031	−0.005	0.068	0.023	−0.015	0.065
Meta	0.031	−0.003	0.065	0.024	−0.013	0.064
Sucre	0.029	−0.01	0.068	0.016	−0.025	0.058
Nariño	0.028	−0.007	0.065	0.018	−0.02	0.058
Magdalena	0.028	−0.006	0.064	0.025	−0.011	0.057
Norte de Santander	0.028	−0.009	0.065	0.017	−0.022	0.059
Tolima	0.023	−0.012	0.061	0.016	−0.022	0.057
Boyacá	0.023	−0.015	0.062	0.017	−0.022	0.061
Risaralda	0.022	−0.013	0.059	0.01	−0.029	0.055
Caldas	0.02	−0.017	0.058	0.008	−0.033	0.054
Cauca	0.018	−0.014	0.052	0.002	−0.036	0.048
Córdoba	0.018	−0.017	0.054	0.008	−0.03	0.048
Santander	0.016	−0.02	0.053	0.017	−0.02	0.054
Bolivar	0.015	−0.02	0.051	0.011	−0.025	0.049
Cundinamarca	0.012	−0.024	0.047	0.009	−0.029	0.047
Atlántico	0.008	−0.026	0.044	0.001	−0.034	0.043
Valle	0.001	−0.034	0.038	0	−0.036	0.04
Antioquia	−0.004	−0.039	0.033	0	−0.035	0.039
Bogotá, D.C.	−0.014	−0.048	0.022	−0.005	−0.04	0.03

Source: Authors' calculations.

7. CONCLUSIONS

Following the seminal proposal by Nishimizu and Page (1982), this study estimates a stochastic translog production frontier to determine the factors driving the total factor productivity change of Colombian regions. After completing the literature review, two situations were found: from the theoretical approach, empirical applications developed diverse proposals to introduce human capital, but none of them included the presence of infrastructures provided by public capital as an additional production factor that could complement existing levels of private investments. From the empirical perspective, we found no applications for Latin-American countries. In light of these gaps, this paper develops an empirical application that considers these two situations.

Two translog frontier production functions were estimated. The original one—formalized in equation (16.6)—considers the standard variables (GDP as output, number of employees and private capital investments as inputs), and the extended model—formalized in equation (16.8)—incorporates an additional input (public capital stock) to the variables included in model (16.6). The data considers Colombian regions for 14 years (1996–2009). These two models were estimated to determine: (1) the extent to which convergence exists among Colombian regions in terms of output growth, (2) the bias (if any) arising from the incomplete definition of the technology in quantifying the total factor productivity and (3) the technological characteristics of public capital as an additional input in the production function.

The overall conclusion is that the impact of the total factor productivity on output growth is more significant in less advanced and smaller regions, a situation that coincides in both the original and the extended estimations (so, $\dot{TFP}_{it}$ rates can help to obtain convergence on output growth). These results confirm that inefficiency in less developed regions is causing strong divergence in terms of growth rates, but advanced regions are unable to generate technological progress. Active plans are therefore needed to stimulate the proper use of available resources in less developed regions. Another possible driver of economic growth is the equilibrium of public and private capital provisions, as clear indications of disequilibria are present in Colombian regions. This situation can cause unavoidable bottlenecks and congestion in the process of generating economic growth from capital stock investments.

One limitation of the study is that the panel data could not be extended to more recent years; we will attempt to overcome this challenge in future applications (this problem was caused by a change in the methodology used to provide the statistical information). Additional extensions might consider the driving factors of economic growth, bias in the levels of technological change, congestion in public infrastructures and the regions' dependence on international markets.

APPENDIX

Table A16.1 Classification of the articles depending on the estimation method

Parametric methods	Non-parametric methods
Wu (2000)	
Han et al. (2002)	
Liao et al. (2007)	
Li and Liu (2011)	
Zhou et al. (2011)	
Özyurt and Guironnet (2011)	
Lee and Cheng (2011)	Salinas-Jiménez (2003)
Ganzhou (2011)	Murillo-Zamorano (2005)
Arazmuradov et al. (2014)	Henderson et al. (2007)
Colino et al. (2014)	Huang and Su (2013)
Brock and Ogloblin (2014)	Li and Lin (2015)
Mamonov and Pestova (2015)	Beltrán-Esteve and Picazo-Tadeo (2017)
Dall'erba and Llamosas-Rosas (2015)	Cheng et al. (2018)
Shao and Lin (2016)	Feng et al. (2018)
Shi et al. (2017)	Dhillon and Vachharajani (2019)
Brock and Oglibin (2018)	Wang et al. (2020)
Chen (2018)	
Wang et al. (2019)	
Liu et al. (2019)	
Li et al. (2019)	
Rao et al (2019)	

Source: The authors.

Table A16.2 Descriptive statistics of Colombian regions

Region	GDP	Employees (thousands)	Capital stock (private)	Capital stock (public)
Antioquia	45,710,261	2.162	10,755,357	19,148,655
Atlántico	14,136,066	0.770	2,998,866	3,919,075
Bogotá, D.C.	85,306,419	2.881	26,260,262	44,233,758
Bolivar	13,138,978	0.659	2,223,661	4,348,458
Boyacá	7,798,209	0.513	890,584	4,426,042
Caldas	5,685,804	0.394	959,161	3,624,509
Cauca	4,558,126	0.525	880,263	5,727,685
Cesar	5,338,740	0.331	435,256	2,087,222
Chocó	1,068,350	0.155	92,407	906,897
Córdoba	5,769,401	0.522	722,709	3,018,289
Cundinamarca	17,023,601	0.923	3,066,468	9,566,996
Huila	5,789,431	0.366	914,049	3,635,081
La Guajira	4,183,240	0.207	387,559	1,875,663
Magdalena	5,806,781	0.409	476,053	2,044,201
Meta	6,137,953	0.302	1,083,762	4,955,448
Nariño	4,865,251	0.669	672,728	2,889,174
N. Santander	5,290,684	0.488	571,689	2,903,137

Source: Authors' calculations after Vallecilla (2011) and *DANE*.

NOTES

1. Although we use the geographical concept of region, these Colombian territorial entities receive the name of *departamentos*. *Departmentos* are country subdivisions and are granted a certain degree of autonomy. Each departmento has a Governor and an Assembly, elected by popular vote for a four-year period.
2. The information was obtained from the World Bank, the IMF, the Fraser Institute and UNESCO.
3. The translog production function was proposed by Christensen, Jorgensen and Lau (1971), and Griliches and Ringstad (1971). Applications in the field of productivity studies are Lau and Brada (1990), Nishimizu and Page (1982) and Young (1995).
4. The monotonicity condition is not fulfilled for some observations of both models, although we can reject a possible model misspecification.
5. The value of the Likelihood-ratio test is 41.90 with a p-value of 0.0000.

REFERENCES

Abramovitz, M. (1986). “Catching up, forging ahead, and falling behind”. *The Journal of Economic History*, 46 (2): 385–406.

Arazmuradov, A., Martini, G., and Scotti, D. (2014). “Determinants of total factor productivity in former Soviet Union economies: A stochastic frontier approach”. *Economic Systems*, 38 (1): 115–135.

Battese, G. E., and Coelli, T. J. (1992). “Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India”. *Journal of Productivity Analysis* 3: 153–169.

Baumol, W. J. (1986). “Productivity growth, convergence, and welfare: What the long-run data show”. *The American Economic Review*, 76 (5): 1072–1085.

Beltrán-Esteve, M. and Picazo-Tadeo, A. J. (2017). “Assessing environmental performance in the European Union: Eco-innovation versus catching-up”. *Energy Policy*, 104: 240–252.

Brock, G. and Ogloblin, C. (2014). “Another look at technical efficiency in American states, 1979–2000”. *The Annals of Regional Science*, 53 (2): 577–590.

Brock, G. and Ogloblin, C. (2018). “Russian 1998–2007 TFP decomposed: Some inspiration emerging from inherited Soviet legacy”. *Economic Change and Restructuring*, 51 (2): 135–151.

Chen, Y.-H. (2018). “Estimating the speeds of long-run technological catch-up and growth of Total Factor Productivity for countries”. *The Journal of International Trade & Economic Development*, 27 (2): 220–234.

Cheng, Z., Li, L., Liu, J., and Zhang, H. (2018). “Total-factor carbon emission efficiency of China’s provincial industrial sector and its dynamic evolution”. *Renewable and Sustainable Energy Reviews*, 94: 330–339.

Coelli, T. J., Rao, D. S. P., O’Donnell, C. J., and Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media.

Colino, A., Benito-Osorio, D., and Rueda-Armengot, C. (2014). “Entrepreneurship culture, total factor productivity growth and technical progress: Patterns of convergence towards the technological frontier”. *Technological Forecasting and Social Change*, 88: 349–359.

Dall’erba, S. and Llamosas-Rosas, I. (2015). “The impact of private, public and human capital on the US states’ economies: Theory, extensions and evidence”. In: *Handbook of Research Methods and Applications in Economic Geography*. Edward Elgar Publishing.

DANE (2020). “*National Region of Statistics*”. Taken from https://www.dane.gov.co (accessed on 09/20/2020).

Dhillon, A. S. and Vachharajani, H. (2019). “Productivity analysis of coal-fired thermal power plants in India using Malmquist index approach”. *Global Business Review*, 20 (4): 1–16.

Farrell, M. J. (1957). “The measurement of productive efficiency”. *Journal of the Royal Statistical Society: Series A General*, 120 (3): 253–281.

Feng, C., Huang, J.-B., and Wang, M. (2018). “Analysis of green total-factor productivity in China’s regional metal industry: A meta-frontier approach”. *Resources Policy*, 58: 219–229.

Ganzhou, L. (2011). “An empirical study on technological progress, technical efficiency change and economic growth in north-east of China”. In: *2011 International Conference on Business Management and Electronic Information*, Volume 2. IEEE, pp. 331–335.

González, J. E., Sierra, L. P., and Mora, J. U. (2022). “Formas de violencia y su relación con la inversión manufacturera departamental en Colombia: 2000-2013”. *Revista Criminalidad*, 64 (1): 1–37.

Hamilton, L. C. (2012). *Statistics with Stata: version 12*. Cengage Learning.

Han, G., Kalirajan, K., and Singh, N. (2002). “Productivity and economic growth in East Asia: innovation, efficiency and accumulation”. *Japan and the World Economy*, 14 (4): 401–424.

Henderson, DJ, Tochkov, K., and Badunenko, O. (2007). “A drive up the capital coast? Contributions to post-reform growth across Chinese provinces”. *Journal of Macroeconomics*, 29 (3): 569–594.

Huang, Z. and Su, X. (2013). “Total factor productivity and decomposition of the output growth sources of the three leading industries in Fujian province: Based on non-parametric production frontier analysis”. *Systems Engineering-Theory & Practice*, 4: 1–9.

Holtz-Eakin D. (1994). “Public-sector capital and the productivity puzzle”. *Review of Economics and Statistics* 6: 12–21.

Juknys, R., Liobikienė, G., and Dagiliūtė, R. (2017). “Sustainability of economic growth and convergence in regions of different developmental stages”. *Sustainable Development*, 25 (4): 276–287.

Lee, Y. H. and Cheng, M. Y. (2011). “Catching-up and technological progress of the Asean economies”. *Pacific Economic Review*, 16 (2): 236–254.

Li, K. and Lin, B. (2015). “Metafrontier energy efficiency with CO2 emissions and its convergence analysis for China”. *Energy Economics*, 48: 230–241.

Li, K.-W. and Liu, T. (2011). “Economic and productivity growth de-composition: An application to post-reform China”. *Economic Modelling*, 28 (1–2): 366–373.

Li, M., Yu, G., and Yang, L. (2019). “Study on total factor productivity of Nanshan district in Shenzhen”. In: *Recent Developments in Intellectual Computing, Communication and Devices*, pp. 35–41. Springer.

Liao, H., Holmes, M., Weyman-Jones, T., and Llewellyn, D. (2007). "Productivity growth of east Asia economies' manufacturing: A decomposition analysis". *The Journal of Development Studies*, 43 (4): 649–674.
Liu, H. and Tsai, H. (2021). "A stochastic frontier approach to assessing Total Factor Productivity change in China's star-rated hotel industry". *Journal of Hospitality & Tourism Research*, 45 (1): 1–24.
Liu, Y., Yao, X., and Wei, T. (2019). "Energy efficiency gap and target setting: A study of information asymmetry between governments and industries in China". *China Economic Review*, 57: 1–10.
Mamonov, M. and Pestova, A. (2015). "The technical efficiency of national economies: Do the institutions, infrastructure and resources rents matter?". *Journal of the New Economic Association*, 27 (3): 44–78.
Montolio, D., and Solé-Ollé, A. (2009). "Road investment and regional productivity growth: The effects of vehicle intensity and congestion". *Papers in Regional Science*, 88 (1), 99–118.
Murillo-Zamorano, L. R. (2005). "The role of energy in productivity growth: A controversial issue?". *The Energy Journal*, 26 (2): 69–88.
Nishimizu, M. and Page, J. M. (1982). "Total factor productivity growth, technological progress and technical efficiency change: Dimensions of productivity change in Yugoslavia, 1965-78". *The Economic Journal*, 92: 920–936.
Özyurt, S. and Guironnet, J.-P. (2011). "Productivity, scale effect and technological catch-up in Chinese regions". *Journal of Chinese Economic and Foreign Trade Studies*, 4 (2): 64–80.
Pereira, A. M., and De Frutos, R. F. (1999). "Public capital accumulation and private sector performance". *Journal of Urban Economics*, 46 (2): 300–322.
Rao, G., Su, B., Li, J., Wang, Y., Zhou, Y., and Wang, Z. (2019). "Carbon sequestration total factor productivity growth and decomposition: A case of the Yangtze river economic belt of China". *Sustainability*, 11 (23): 1–28.
Sala i Martín, X. (2000). *Notes on economic growth*. Antoni Bosch Editor.
Salinas-Jiménez, M. M. (2003). "Technological change, efficiency gains and capital accumulation in labor productivity growth and convergence: An application to the Spanish regions". *Applied Economics*, 35 (17): 1839–1851.
Shao, B. B. and Lin, W. T. (2016). "Assessing output performance of information technology service industries: Productivity, innovation and catch-up". *International Journal of Production Economics*, 172: 43–53.
Shi, C., Jie, C., Shi, F., and Wen, Y. (2017). "Estimation of the technical efficiency and the decomposition of TFP of Chinese forestry: Basing on sfa-Malmquist method". *Linye Kexue / Scientia Silvae Sinicae*, 53 (12): 126–135.
Solow, R. M. (1957). "Technical change and the aggregate production function". *The Review of Economics and Statistics*, 39 (3): 312–320.
Vallecilla, J. (2011). "GDP, investment and regional capital stock in Colombia 1975-2009". *Region Es*, 6 (2): 43–84.
Wang, C.-N., Tibo, H., and Nguyen, H. A. (2020). "Malmquist productivity analysis of top global automobile manufacturers". *Mathematics*, 8 (4): 1–21.
Wang, Y., Hao, C., and Zhang, Z. (2019). "Influence of environmental regulation on the growth of total factor productivity - An empirical research based on China's provincial panel data". *Applied Ecology and Environmental Research*, 17 (4): 8417–8434.
Wu, Y. (2000). "Is China's economic growth sustainable? To productivity analysis". *China Economic Review*, 11 (3): 278–296.
Zhou, X., Li, K.-W., and Li, Q. (2011). "An analysis on technical efficiency in post-reform China". *China Economic Review*, 22 (3): 357–372.

17. Analysing public sector efficiency of the Indian States

Ranjan Kumar Mohanty, N R Bhanumurthy, and Biresh K. Sahoo

1. INTRODUCTION

Every government is mandated to provide various public goods and services for maximizing its citizens' welfare. However, many governments, especially in the developing and least developed countries, are stressed with severe budget constraints (high fiscal deficit and public debt). Compared to national governments, local governments often face resource constraints while delivering public goods. Effective allocation of scarce resources could solve the problem of resource scarcity (to some extent), or conversely, the public sector can use their existing resources in a more efficient manner to provide services to their citizens. With the pressure on governments increasing due to citizens' high expectations from the government (transparency of government practices), this exerts higher pressure on the policymakers to utilize the existing resources more efficiently. Therefore, the efficient management of local governments' scarce resources has been a topic of high interest to researchers and policymakers.

Few empirical studies have evaluated local governments' performances in several countries. These studies are confined mostly to advanced, OECD or European countries (see, for example, Kalb, 2010, and Geys et al., 2010 for Germany; Worthington, 2000 for Australia; De Borger & Kerstens, 1996 for Belgium; Doumpos & Cohen, 2014 for Greece; Lo Storto, 2016 for Italy; Balaguer-Coll et al., 2010, and Pérez-López et al., 2015 for Spain; Sampaio & Stosic, 2005 for Brazil; Afonso & Fernandes, 2008 for Portugal; Revelli, 2010 for the United Kingdom; Grossman et al., 1999 for the United States; Loikkanen & Susiluoto, 2005 for Finland; Kalseth & Rattso, 1995 for Norway; among others). These studies have used both the non-parametric approaches such as DEA, Free Disposal Hull (FDH) and so on (see De Borger & Kerstens, 1996; Afonso & Fernandes, 2006; Lo Storto, 2016; Mohanty & Bhanumurthy, 2020 and others) and parametric methods such as Stochastic Frontier Analysis (SFA), Ordinary Least Square (OLS), panel regressions and so on (see, for example, Grossman et al., 1999; Worthington, 2000; Kalb, 2010; Kalseth & Rattso, 1995; Revelli, 2010; among others) to analyze local governments' efficiency. The key findings from these studies reflect that the efficiency of public sector spending could be improved by focusing on various socio-economic and environmental factors.[1]

For an emerging economy such as India, public spending in education, health, infrastructure, law and order, and so on, is crucial for its rapid and sustainable development. With a vast population, India has a wide regional disparity because of its geographic, economic and demographic characteristics. The divergence in economic growth and human development across the Indian States has been posing many challenges for governments. India's current resource allocation policy (through both horizontal and vertical equalization principles) has not impacted the overall development outcomes to address such regional divergences. Adoption of the Fiscal Responsibility and Budget Management (FRBM) Act has also added a layer to the budget constraints faced by various State governments. To make public finances sustainable, the efficiency of public spending could play a crucial role in a resource-scarce economy such as that of India.[2] To achieve the medium-term target of a US$5 trillion Indian economy, a comprehensive analysis of state-wise public sector performance becomes crucial. Public spending as a percent of Gross State Domestic Product (GSDP) varied between 11 and 74 percent across States in 2018–19.[3] Whether or not such diverse public spending is being made efficiently is an important issue in providing the best possible public service outcomes at an optimal level. Hence, small changes in the efficiency of public spending could have a significant impact on the overall budget as well as on economic development. The issue of efficiency in public sector spending becomes much more crucial at the State level, as the budget constraints, citizens' demands and absorptive capacities are quite diverse. Therefore, some of the research questions pertinent at this juncture are: do the Indian States utilize their public sector spending efficiently? Do the Indian States perform differently across different public sectors?

Until recently, there has been little discussion on the efficiency of the public sector in the Indian States, especially at the individual sector level, and a large part of the discussion was on outlays and not much on outputs/outcomes.[4] Hence, there is a need to understand the outlays–outputs framework to reprioritize and recalibrate public expenditures, and this could be done by assessing public sector spending efficiency. In this chapter, this is done at the Indian States level as a State is the major implementing agency of most public sector programs. Here, the chapter uses the non-parametric Data Envelopment Analysis (DEA) method (Charnes et al., 1978; Banker et al., 1984). Our results are expected to help in understanding whether some regions lag behind others due to lack of resources, inefficiency in using resources or a combination of both. The study finds a significant difference in the efficiency of public sectors, as well as scope for improvement across the Indian States. Most of the southern and western States have better efficiency levels compared to other regions.

By undertaking the study on India, this study extends the nascent literature in the following directions. To the best of our knowledge, it is the first of its kind to evaluate the efficiency of eleven major public sector spending (individually) of 29 Indian States.[5] Such an analysis will help policymakers to implement sector-specific policies. Then, along with a separate efficiency analysis of eleven public sectors, it also constructs three aggregate public-sector efficiency indices, that is, the social sector efficiency index, the economic and general sector efficiency index and the aggregate public sector performance index. This type of efficiency analysis has not been attempted in earlier Indian studies. The aggregate efficiency score could be very useful for comparing public sector performance across the Indian States.

The broad structure of this chapter is as follows. A brief literature review on this issue is presented in Section 2. The data and methodology used for the empirical analysis are discussed in Section 3. Section 4 discusses the estimated DEA results in detail. Section 5 summarizes and discusses the possible policy implications.

2. THE LITERATURE REVIEW

This section highlights some of the significant studies on public sector efficiency, especially at the local/regional government level. An extensive and comprehensive existing literature review on local governments' efficiency is presented in Worthington and Dollery (2000), Afonso and Fernandes (2008), Kalb et al. (2012), Da Cruz and Marques (2014) and Narbon-Perpinaa and De Witte (2018). Many researchers have contributed extensively to this public sector efficiency literature (Afonso et al., 2005; Afonso and Aubyn, 2005; Gupta and Verhoeven, 2001).

Antonelli and De Bonis (2019) examine the public spending efficiency of 22 European countries and suggest that the existing output can be achieved with a 20 percent (on average) reduction in public expenditure. The studies point out that high efficiency is associated with higher GDP, higher education level, lower corruption and a relatively smaller population. Using the case of Norwegian local governments, Borge et al. (2008) find that high fiscal capacity, a centralized top-down budgetary procedure and a high degree of party fragmentation decreases while democratic participation increases the efficiency of public service provisions. Afonso and Fernandes (2008) examine the efficiency of 278 Portuguese municipal governments and find that education level, per capita purchasing power and distance to the district's capital contributed positively to increased efficiency. Afonso et al. (2010) measure the public sector efficiency for ten new member states of the European Union and emerging markets. The study finds that education level, higher income, property rights security and civil service competence help prevent public sector inefficiencies.

The efficiency of health care, school education, general public services and so on, has been examined for OECD regions (Afonso and Aubyn, 2005; Sutherland et al., 2010; Joumard et al., 2010; Dutu and Sicari, 2016). Adam, Delis and Kammas (2018) find a wide variation in the efficiency of education, health, social security and welfare, general public services and economic affairs among OECD countries. The study finds that various factors such as greater fiscal decentralization, domestic participation, strong political leadership, right-wing government and so on are the major determinants of efficiency. Afonso et al. (2005) analyze the public sector performance in 23 industrialized OECD countries for 1990 and 2000 and suggest a significant difference in public sector performances, indicating potential for expenditure saving in many countries.

Feeny and Rogres (2008) use stochastic frontier analysis (SFA) to estimate the efficiency of public sector expenditures and foreign aid in 37 Small Island Developing States (SIDS). The study finds governance has a positive impact while there is a mixed result of public sector efficiency and foreign aid in achieving social sector outcomes. Using a sample of 64 developed and developing countries, Angelopoulos et al. (2008) find the relation between fiscal size and economic growth depends on the size–efficiency mix of public sectors. Hwang and Akdede (2011) find that quality of governance has a positive impact on public sector efficiency in administration, infrastructure and stability. Hauner (2008) finds that the differences in public sector efficiency in Russia's regions were explained by the level of spending, governance quality, per capita income, democratic control and the share of federal transfers.

Gupta and Verhoeven (2001) find that African countries were less efficient (on average) than the Western Hemisphere and the Asian countries in providing health and education services during 1984–95. Chan and Karim (2012) find that East Asian countries were relatively less efficient in public spending for promoting income equality during 2000–07, indicating that political stability and financial freedom have contributed positively to public spending efficiency. Ouertani, Naifar and Haddad (2018), using a sample from Saudi Arabia over the period 1988–2013, show that public sector performance on health, education and infrastructure can be improved without increasing the existing spending.

Few studies are carried out in the Indian context. Saxena et al. (2018) analyze the relationship between public infrastructure investment and economic growth for 28 Indian States (infrastructure spending efficiency) over 2010–13. Yadava and Neog (2019) examine the performance of 19 Indian States for the period of 2006–15 using seven indicators and find that states could decrease public spending by 57.88 percent and still achieve the current output. In an earlier attempt, Mohanty and Bhanumurthy (2020) assess the social sector's public expenditure efficiencies, especially health and education, among the major Indian States. These Indian studies cover limited public sector indicators.

The above literature review points out that the assessment of most of the public expenditure efficiency literature is confined to advanced or European countries. There are few studies addressing the Asian region, especially India. Most of the studies have measured individual public spending efficiency (usually education, health and so on), while very few have focused on evaluating the public sector efficiency. This chapter tries to bridge this gap by analyzing a wide range of public sector spending across the Indian States during 2005–18 (a more extended period). The next section presents the data and methodology used in the study.

3. THE DATA AND METHODOLOGY

3.1 The Data

The public sector efficiency of all the 29 Indian States is selected for the empirical analysis.[6] The study period is 2005–18,[7] and thus, the efficiency is calculated for two different periods, that is, 2005–06 and 2018–19. We have carried out the efficiency analysis for 11 major public sector expenditures: education, health, water supply and sanitation, information and broadcasting, social welfare and nutrition, rural development, irrigation and flood control, energy, road transport, public works and police.[8] These services cover social services, economic services and general services of the respective States' budgets.[9]

The study has collected many relevant variables on both outlays and outputs from several sources, and they are as follows. All public expenditure related variables, such as the public expenditure on education, health, water supply and sanitation, information and broadcasting, social welfare and nutrition, rural development, irrigation and flood control, energy, roads transport, public works and police, are assembled from the NIPFP Databank and the 'State Finances: A study of Budget', Reserve Bank of India (RBI). Data on Gross State Domestic Product (GSDP) are collected from National Accounts Statistics, Central Statistics Office (CSO). Information on Population is acquired from the report 'Population Projections for India and States 2001-2026', Office of The Registrar General and Census Commissioner (ORGCC), Government of India. Many variables such as Gross State Value Added (GSVA) by agriculture, GSVA by services, GSVA by construction, State-wise telephones per 100 population, the length of State highways, the length of national highways, the total length of roads, gross irrigated area, gross shown area, per capita availability of power and installed capacity of power are sourced from the Handbook of Statistics on Indian States (HSIS), RBI.

Similarly, the gross enrolment ratio for higher education and various health infrastructures are gathered from the EPWRF India Time Series, EPW Research Foundation. Data for wasting and stunting are collected from the reports of the National Family Health Survey (NFHS)-3 and NFHS-4. Access to safe drinking water and latrines are assembled from the report 'Drinking

Water, Sanitation, Hygiene, and Housing Condition in India,' Ministry of Statistics and Programme Implementation (MOSPI), Government of India, and also from the Indiastat.com. Data on the length of rural roads and road accidents are gathered from various reports of the 'Basic road statistics of India' and 'Road accidents in India,' respectively, Ministry of Road Transport and Highways, Government of India. Further, the reading performance indicator is obtained from the Annual Status of Education Report (ASER) report of 2005 and 2018, while the infant mortality rate (IMR) is sourced from Sample Registration System (SRS) Bulletins and ORGCC (Government of India). Finally, data on the crime disposal rate and charge sheeting rate are obtained from various issues of the crime reports in India, National Crime Records Bureau (NCRB), and the yield on Rabi food grains are sourced from the report of Agricultural Statistics at a Glance 2018, Department of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India. The methodology used in the study is addressed in the next section.

3.2 Data Envelopment Analysis Methodology

There are two alternative approaches followed to evaluate the efficiency of firms: the econometric approach and the data envelopment analysis (DEA) approach. The advantage of the econometric approach is that it allows for formal statistical testing of hypotheses and the construction of confidence intervals. However, the problem underlying this approach is that it is parametric and can confound the effects of misspecification of functional form with scale economies; and further, *flexible* functional forms are susceptible to multicollinearity, and theoretical restrictions may be violated. The DEA approach, however, has the advantage of both being non-parametric, which means less susceptible to specification error, and being able to accommodate multiple inputs and multiple outputs (Sahoo & Gstach, 2011).[10] The current study has therefore used the DEA approach for measuring the public sector efficiency.

When a DMU produces the maximum level of output(s) with a given amount of input(s), it is called efficient. Although there are different oriented models for measuring efficiency, we have chosen the most appropriate output-oriented model in our study. An input-oriented model was not considered appropriate because a reduction of public spending would not be an ideal policy option for the Indian States as the States have already been experiencing a very low share of public sector spending in their GSDP.

Consider the evaluation of a sample of n decision making units, DMUs (States). Each State j ($j = 1, \ldots, n$) is assumed to consume an m-input vector $x_j = \left(x_{ij}\right)_{i=1,\ldots,m} \in R_+^m$ to produce an s-output vector $y_j = \left(y_{rj}\right)_{r=1,\ldots,s} \in R_+^s$. The relative output efficiency (E_o) of any State o ($o = 1, \ldots, n$) is a comparison between its *actual* outputs produced and the *best-practice* (*maximal*) outputs that would have been produced from its current level of inputs. The idea underlying this output efficiency measure is to reallocate the input vector x_o over all the n States, and then run them with the respective intensities $\lambda_1, \lambda_2, \ldots, \lambda_n$ to expand the observed output vector y_o by an expansion factor θ, which can be computed from the following linear program (Banker et al., 1984):

$$\left(E_o\right)^{-1} = \max_{\theta,\lambda_j} \theta \tag{17.1}$$

s.t.

$$\sum\nolimits_{j=1}^{n} x_{ij}\lambda_j \le x_{io} \; \left(\forall i\right), \tag{17.1.1}$$

$$-\sum_{j=1}^{n} y_{rj}\lambda_j + \theta y_{ro} \leq 0 \left(\forall r\right), \tag{17.1.2}$$

$$\sum_{j=1}^{n} \lambda_j = 1, \tag{17.1.3}$$

$$\theta \geq 0, \lambda_j \geq 0 \left(\forall j\right). \tag{17.1.4}$$

The efficiency evaluation model (17.1) is based on two assumptions: (a) the firms $(x_j, y_j)_{j=1,\ldots,n}$ are run with variable returns to scale (which is ensured by the use of convexity constraint (17.1.3)) and (b) the inputs and outputs are freely disposable (which is ensured by the use of inequality signs in the input and output constraints, that is, (17.1.1) and (17.1.2)). The model (17.1) determines State o's maximal output as $(E_o)^{-1}y_o$ from its current input vector x_o using the best practice technologies. State o's output efficiency is then computed as the ratio of its actual output vector y_o to its potential output vector x_o, that is, $E_o = 1/\theta^*$, which lies between 0 and 1, where θ^* is the optimal solution obtained from the model (17.1).

The empirical results of the output-oriented approach are presented and discussed in the next section.

4. EMPIRICAL ANALYSIS

As discussed earlier, the efficiency levels of 11 major public sector spending activities are examined in this chapter using the output-oriented DEA approach among 29 States. These public services broadly fall under social services, economic services and general services in each of these States' budgets. As the collection of public finance data on all the public services at the state level is tedious, especially at the local level, this chapter is limited to 11 major sectors. Variable returns to scale are assumed due to the wide variation among States. The results are discussed below.

4.1 Social Services

Social sector spending is necessary and an essential source of human development. In this section, within the social sector, the public sector spending efficiencies of education, health, water supply and sanitation, social welfare and nutrition and information and broadcasting are discussed as follows.

4.1.1 Education

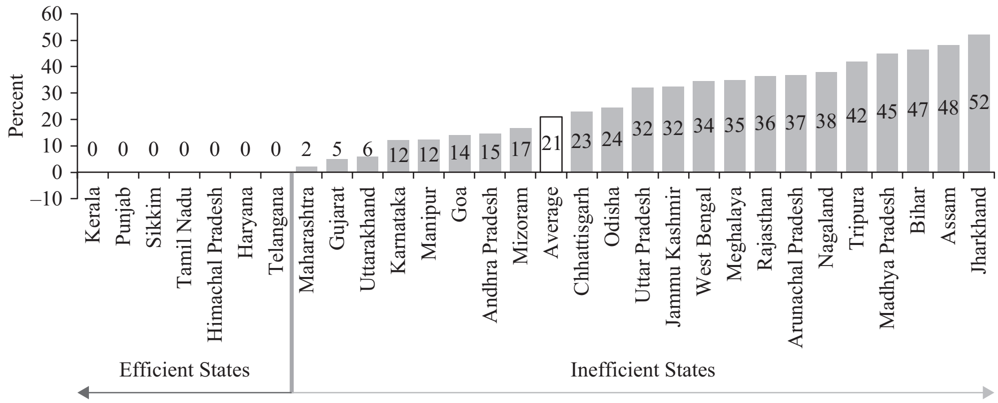
We have used two outputs, that is, the gross enrollment ratio for higher education and the percentage of children who can read Standard-II level at Standard-V (learning level), and one input, that is, education expenditure to GSDP ratio for measuring the efficiency of education. It is to be noted that the data for education spending are averaged over two periods (2004–2005 for 2005, and 2017–2018 for 2018). The estimated DEA results using the output-oriented approach for education are given in Table 17.1 and Figure 17.1.

The results suggest that, in 2018 (Table 17.1), out of 29 States, seven States were labeled as efficient, namely Kerala, Punjab, Sikkim, Tamil Nadu, Haryana, Himachal Pradesh and

Table 17.1 Education efficiency of Indian States (output oriented)

	Output Oriented_2005		Output Oriented_2018		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Andhra Pradesh	0.90	12	0.85	14	Haryana, Kerala, Tamil Nadu.
Arunachal Pradesh	0.76	22	0.63	23	Himachal Pradesh, Sikkim.
Assam	0.61	26	0.52	28	Kerala.
Bihar	0.81	20	0.53	27	Kerala.
Chhattisgarh	0.95	10	0.77	16	Kerala.
Goa	0.95	9	0.86	13	Kerala.
Gujarat	0.82	19	0.95	9	Haryana, Telangana.
Haryana	1.00	1	1.00	1	Har.
Himachal Pradesh	1.00	1	1.00	1	Himachal Pradesh.
Jammu Kashmir	0.78	21	0.68	19	Himachal Pradesh, Sikkim.
Jharkhand	0.82	18	0.48	29	Himachal Pradesh, Kerala, Sikkim.
Karnataka	0.87	16	0.88	11	Haryana, Telangana.
Kerala	1.00	1	1.00	1	Kerala.
Madhya Pradesh	0.89	15	0.55	26	Himachal Pradesh, Kerala, Sikkim.
Maharashtra	1.00	1	0.98	8	Haryana, Kerala, Tamil Nadu.
Manipur	0.86	17	0.88	12	Himachal Pradesh, Kerala.
Meghalaya	1.00	1	0.65	21	Himachal Pradesh, Sikkim.
Mizoram	-	-	0.83	15	Kerala.
Nagaland	0.90	13	0.62	24	Kerala.
Odisha	0.75	23	0.76	17	Kerala.
Punjab	0.90	14	1.00	1	Punjab.
Rajasthan	0.74	24	0.64	22	Kerala.
Sikkim	-	-	1.00	1	Sikkim.
Tamil Nadu	1.00	1	1.00	1	Tamil Nadu.
Telangana	-	-	1.00	1	Telangana.
Tripura	0.92	11	0.58	25	Kerala.
Uttar Pradesh	0.68	25	0.68	18	Himachal Pradesh, Kerala, Sikkim.
Uttarakhand	1.00	1	0.94	10	Himachal Pradesh, Kerala, Sikkim.
West Bengal	0.96	8	0.66	20	Kerala.
Average	**0.88**		**0.79**		

Source: Authors' calculation.



Source: Authors' estimation.

Figure 17.1 Education inefficiency, 2018

Telangana. Among these States, only Kerala, Haryana, Himachal Pradesh and Tamil Nadu are consistently efficient both in 2005 and 2018. These efficient States have achieved a higher enrollment ratio and also a higher percentage of learning level using a smaller proportion of resources than the national average. The output-oriented results show that the bottom three States were Bihar, Assam and Jharkhand in 2018. The average education efficiency score is 0.79, implying, on average, that States are producing 21 percent fewer outputs than they should if they were efficient. For example, the efficiency score of Odisha is 0.76, which implies that only 76 percent of outputs is produced with the existing resources (24 percent output can be enhanced). Nearly half of the States are utilizing their public spending on education inefficiently above the all-State average (Figure 17.1). Figure 17.1 also suggests that many Indian States could enhance their education output with the existing level of education spending by following the best practice (Karnataka and Manipur-12 percent, Goa-14 percent, Andhra Pradesh-15 percent, Mizoram-17 percent, Chhattisgarh-23 percent, Odisha-24 percent, Uttar Pradesh, and Jammu and Kashmir-32 percent, West Bengal-34 percent, Meghalaya-35 percent, Rajasthan-36 percent, Arunachal Pradesh-37 percent, Nagaland-38 percent, Tripura-42 percent, Bihar-47 percent, Assam-48 percent, Jharkhand-52 percent).

4.1.2 Health

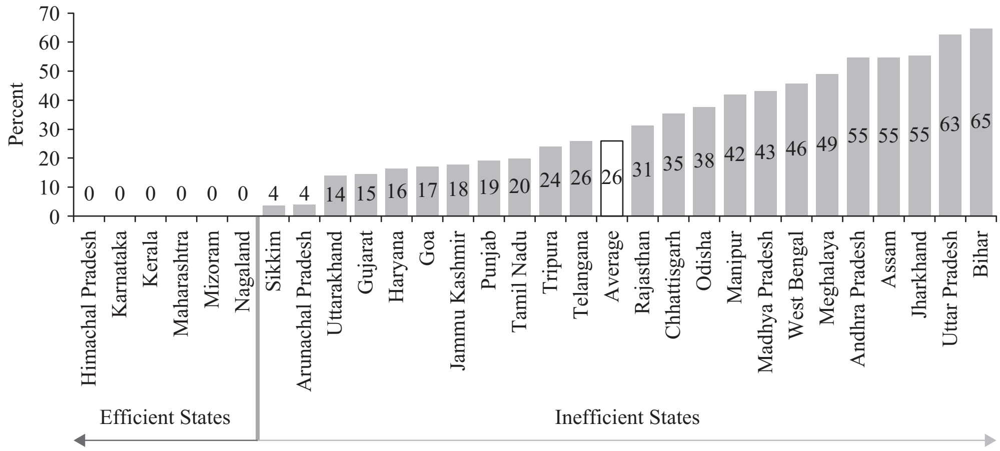
For health, efficiency is analyzed using two outputs and one input. The two outputs are the infant mortality rate (IMR) and health care infrastructure.[11] The health expenditure to GSDP ratio[12] is used as an input for measuring health efficiency. The DEA techniques imply that outputs are measured in such a way that 'more is better'. Here, the IMR refers to the (number of children who died before 12 months)/(number of children born)×1000. Therefore, we have calculated an 'Infant Survival Rate (ISR)' as follows. ISR[13] = (1000−IMR)/ IMR.

Finally, ISR and health infrastructure are used as final outputs in the DEA approach. Health spending data are averaged over two periods, that is, 2004−2005 for 2005 and 2017−2018 for 2018. The results of the health sector efficiency scores using an output-oriented approach are given in Table 17.2 and Figure 17.2.

Table 17.2 Health efficiency of Indian States (output oriented)

	Output Oriented_2005		Output Oriented_2018		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Andhra Pradesh	0.57	25	0.45	25	Himachal Pradesh, Karnataka, Kerala.
Arunachal Pradesh	0.97	9	0.96	8	Mizoram.
Assam	0.80	13	0.45	26	Himachal Pradesh, Mizoram.
Bihar	0.42	28	0.35	29	Himachal Pradesh, Kerala, Mizoram.
Chhattisgarh	1.00	1	0.65	19	Himachal Pradesh, Karnataka.
Goa	0.85	11	0.83	12	Kerala, Nagaland.
Gujarat	1.00	1	0.85	10	Himachal Pradesh, Karnataka.
Haryana	1.00	1	0.84	11	Karnataka, Maharashtra.
Himachal Pradesh	1.00	1	1.00	1	Himachal Pradesh.
Jammu Kashmir	0.54	26	0.82	13	Himachal Pradesh, Karnataka, Kerala.
Jharkhand	0.62	21	0.45	27	Himachal Pradesh, Karnataka, Kerala.
Karnataka	0.86	10	1.00	1	Karnataka.
Kerala	1.00	1	1.00	1	Kerala.
Madhya Pradesh	0.60	23	0.57	22	Himachal Pradesh, Karnataka.
Maharashtra	1.00	1	1.00	1	Maharashtra.
Manipur	1.00	1	0.58	21	Himachal Pradesh, Kerala, Mizoram.
Meghalaya	0.58	24	0.51	24	Himachal Pradesh, Mizoram.
Mizoram	1.00	1	1.00	1	Mizoram.
Nagaland	0.81	12	1.00	1	Nagaland.
Odisha	0.79	14	0.62	20	Himachal Pradesh, Karnataka.
Punjab	0.66	19	0.81	14	Karnataka, Kerala, Maharashtra.
Rajasthan	0.74	16	0.69	18	Himachal Pradesh, Karnataka.
Sikkim	0.67	18	0.96	7	Himachal Pradesh, Kerala, Mizoram.
Tamil Nadu	0.78	15	0.80	15	Himachal Pradesh, Karnataka, Kerala.
Telangana			0.74	17	Himachal Pradesh, Karnataka, Kerala.
Tripura	0.61	22	0.76	16	Himachal Pradesh, Mizoram.
Uttar Pradesh	0.48	27	0.37	28	Himachal Pradesh, Karnataka, Kerala.
Uttarakhand	0.69	17	0.86	9	Himachal Pradesh, Karnataka.
West Bengal	0.65	20	0.54	23	Himachal Pradesh, Karnataka, Kerala.
Average	0.77		0.74		

Source: Authors' calculation.



Source: Authors' calculation.

Figure 17.2 Health inefficiency, 2018

It is observed from Table 17.2 that Himachal Pradesh, Kerala, Karnataka, Maharashtra, Mizoram and Nagaland were the most efficient States, while Jharkhand, Uttar Pradesh and Bihar were the least efficient States in 2018. The lowest IMRs are found in Kerala, Mizoram and Nagaland, while Karnataka and Maharashtra are among the lowest share of health spending to GDP with a better health outcome. The number of health infrastructures in Himachal Pradesh is better compared to other States. The output efficient score of all States is 0.74. This implies that all States, on average, produce about 26 percent fewer outputs using the same inputs. On average, health efficiency has marginally declined from 0.77 in 2005 to 0.74 in 2018. Figure 17.2 demonstrates the output inefficiency of the remaining States (for example, in 2018; Gujarat, Uttarakhand, Haryana, Punjab, Goa and Jammu and Kashmir-<20 percent; Tamil Nadu, Telangana and Tripura-20 to 30 percent, Rajasthan, Chhattisgarh and Odisha-30 to<40 percent; Manipur, Madhya Pradesh, West Bengal and Meghalaya-40 to 50 percent; Andhra Pradesh, Jharkhand and Assam-55 percent, Uttar Pradesh and Bihar-63 to 65 percent). Four States, Kerala, Himachal Pradesh, Maharashtra and Mizoram, are consistently performing better both in 2005 and 2018. The health efficiency of 12 States is below the all-State average (Figure 17.2). There is the possibility of higher health output using existing resources.

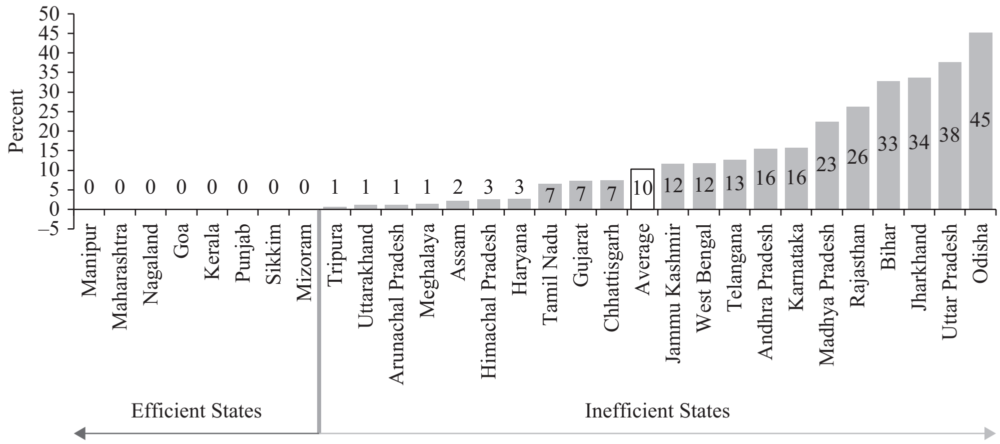
4.1.3 Water supply and sanitation

For measuring efficiency in water supply and sanitation, we have used the budgetary spending on water supply and sanitation to GSDP ratio as an input, while 'percentage of households having access to piped water supply and public tap (safe drinking water)' and 'percentage of households having access to latrine' are the two outputs in the DEA set-up.[14] We have followed a similar strategy of averaging two periods of public expenditure on water supply and sanitation (2004–2005 for 2005, and 2017–2018 for 2018). The results of output-oriented efficiency scores are displayed in Table 17.3 and Figure 17.3.

Table 17.3 Water supply and sanitation efficiency of Indian States (output oriented)

	Output Oriented_2005		Output Oriented_2018		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Andhra Pradesh	0.82	20	0.84	22	Kerala, Punjab, Sikkim.
Arunachal Pradesh	0.89	10	0.99	11	Mizoram.
Assam	0.89	11	0.98	13	Mizoram, Sikkim.
Bihar	0.89	13	0.67	26	Mizoram, Sikkim.
Chhattisgarh	0.72	23	0.93	18	Kerala, Sikkim.
Goa	0.88	15	1.00	4	Goa.
Gujarat	0.86	19	0.93	17	Maharashtra, Punjab, Sikkim.
Haryana	0.88	14	0.97	15	Kerala, Sikkim.
Himachal Pradesh	0.91	9	0.97	14	Mizoram, Sikkim.
Jammu Kashmir	0.81	22	0.88	19	Mizoram, Sikkim.
Jharkhand	0.44	28	0.66	27	Mizoram, Sikkim.
Karnataka	0.87	18	0.84	23	Maharashtra, Punjab, Sikkim.
Kerala	1.00	1	1.00	4	Kerala.
Madhya Pradesh	0.70	24	0.78	24	Mizoram, Sikkim.
Maharashtra	0.82	21	1.00	2	Maharashtra.
Manipur	0.96	6	1.00	1	Mizoram, Sikkim.
Meghalaya	0.66	27	0.99	12	Mizoram, Sikkim.
Mizoram	1.00	1	1.00	8	Mizoram.
Nagaland	0.87	17	1.00	2	Mizoram, Sikkim.
Odisha	0.66	26	0.55	29	Mizoram, Sikkim.
Punjab	1.00	1	1.00	4	Punjab.
Rajasthan	0.70	25	0.74	25	Mizoram, Sikkim.
Sikkim	0.93	8	1.00	4	Sikkim.
Tamil Nadu	0.88	16	0.93	16	Goa, Maharashtra.
Telangana			0.87	21	Kerala, Sikkim.
Tripura	1.00	1	0.99	9	Mizoram, Sikkim.
Uttar Pradesh	0.94	7	0.62	28	Kerala, Sikkim.
Uttarakhand	0.89	12	0.99	10	Kerala, Punjab, Sikkim.
West Bengal	1.00	1	0.88	20	Kerala, Sikkim.
Average	**0.85**		**0.90**		

Source: Authors' calculation.



Source: Authors' calculation.

Figure 17.3 Water supply and sanitation inefficiency, 2018

As can be seen, in 2018, eight States, including Maharashtra, Manipur, Nagaland, Goa, Kerala, Punjab, Sikkim and Tripura, were efficient. The other States, Tripura, Uttarakhand, Arunachal Pradesh, Meghalaya, Assam, Himachal Pradesh and Haryana, were also very close to the efficiency frontier. This shows that all the north-eastern States are utilizing their water supply and sanitation resources very efficiently. These north-eastern States have nearly 100 percent access to latrines, with relatively better access to a piped water supply. Most of these efficient States are achieving a better outcome using a very low share of public spending. On average, the all-State output efficiency score is 0.9, implying they have scope for improvement of an additional 10 percent of their output with the existing resources. The bottom four States are Bihar, Jharkhand, Uttar Pradesh and Odisha. There is scope among many States to enhance their output with the existing resources (Figure 17.3). This demonstrates that the output inefficiencies of Jammu Kashmir, West Bengal, Telangana, Andhra Pradesh and Karnataka are 12 to 16 percent, while Madhya Pradesh, Rajasthan, Bihar, Jharkhand, Uttar Pradesh and Odisha are under the 23 to 45 percent categories. Overall, the output efficiencies of all States have increased from 0.85 to 0.9 in 2005 and 2018, respectively.

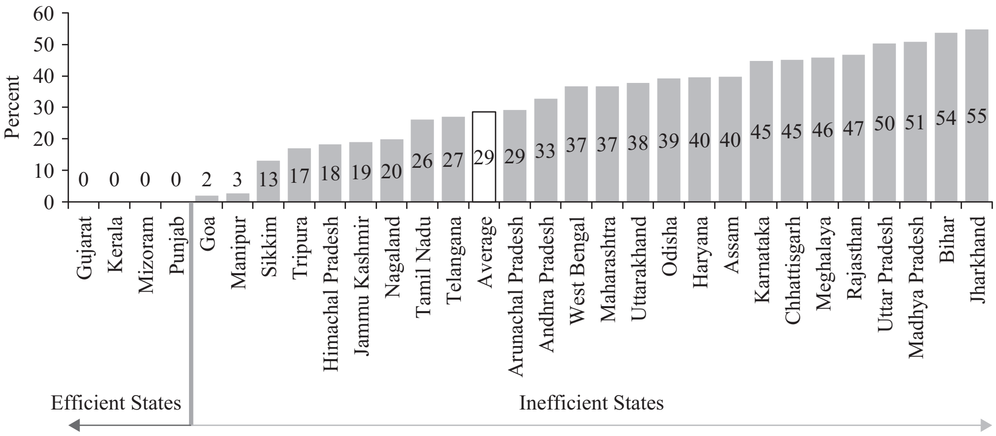
4.1.4 Social welfare and nutrition

Appropriate output data for social welfare and nutrition are not easily available State-wise in India. We have chosen two major pieces of nutrition deficiency-related information, that is, 'wasting' and 'stunting' as two outputs,[15] and public spending on social welfare and nutrition is chosen as an input in the efficiency analysis. The information on wasting and stunting is available in NFHS-3 (conducted in 2005) and NFHS-4 (in 2015). Therefore, we have estimated efficiency for these two years. The results are shown in Table 17.4 and Figure 17.4.

Table 17.4 Social welfare and nutrition efficiency of Indian States (output oriented)

	Output Oriented_2005		Output Oriented_2015		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Andhra Pradesh	0.78	10	0.67	15	Kerala, Mizoram.
Arunachal Pradesh	0.70	12	0.71	14	Kerala, Mizoram.
Assam	0.71	11	0.60	21	Kerala, Mizoram.
Bihar	0.48	27	0.46	28	Kerala, Mizoram.
Chhattisgarh	0.56	24	0.55	23	Kerala, Mizoram.
Goa	1.00	1	0.98	5	Kerala.
Gujarat	0.59	23	1.00	1	Gujarat.
Haryana	0.62	22	0.60	20	Kerala, Mizoram.
Himachal Pradesh	0.69	14	0.82	9	Kerala, Mizoram, Punjab.
Jammu Kashmir	0.80	7	0.81	10	Kerala, Mizoram.
Jharkhand	0.49	25	0.45	29	Kerala, Mizoram.
Karnataka	0.66	16	0.55	22	Kerala, Mizoram.
Kerala	1.00	1	1.00	1	Kerala.
Madhya Pradesh	0.49	26	0.49	27	Kerala, Mizoram.
Maharashtra	0.66	17	0.63	17	Kerala, Punjab.
Manipur	1.00	1	0.97	6	Mizoram, Punjab.
Meghalaya	0.46	28	0.54	24	Kerala, Mizoram.
Mizoram	1.00	1	1.00	1	Mizoram.
Nagaland	0.79	9	0.80	11	Kerala, Mizoram.
Odisha	0.62	21	0.61	19	Kerala, Mizoram.
Punjab	1.00	1	1.00	1	Punjab.
Rajasthan	0.62	20	0.53	25	Kerala, Mizoram.
Sikkim	0.93	6	0.87	7	Kerala, Mizoram, Punjab.
Tamil Nadu	0.79	8	0.74	12	Kerala, Mizoram.
Telangana			0.73	13	Kerala, Mizoram.
Tripura	0.69	13	0.83	8	Kerala, Mizoram.
Uttar Pradesh	0.63	19	0.50	26	Kerala, Mizoram.
Uttarakhand	0.64	18	0.62	18	Kerala, Mizoram.
West Bengal	0.67	15	0.63	16	Kerala, Mizoram.
Average	**0.72**		**0.71**		

Source: Authors' calculation.



Source: Authors' calculation.

Figure 17.4 Social welfare and nutrition inefficiency, 2015

The results show that, out of 29 States, only Gujarat, Kerala, Mizoram and Punjab (four States) were efficient in 2015 and the remaining States had scope for improvement. Bihar and Jharkhand are listed in the bottom category, and they are producing only 44 to 45 percent of their potential output using the existing resources. On average, all States produce about 29 percent (efficiency score 0.71) fewer outputs than they should if they were efficient. The efficiency score of nearly 16 States is far below the average score of all States. It can be seen from Figure 17.4 that many States are producing less (inefficient) with their existing resources (Sikkim, Tripura, Himachal Pradesh, Jammu Kashmir and Nagaland-13 to 20 percent; Tamil Nadu, Telangana, Arunachal Pradesh, Andhra Pradesh, West Bengal, Maharashtra, Uttarakhand, Odisha, Haryana and Assam-21 to 40 percent; Karnataka, Chhattisgarh, Meghalaya, Rajasthan, Uttar Pradesh, Madhya Pradesh, Bihar and Jharkhand-41 to 55 percent). However, the overall output efficiency scores of social welfare and nutrition have more or less remained the same from 2005 to 2015.

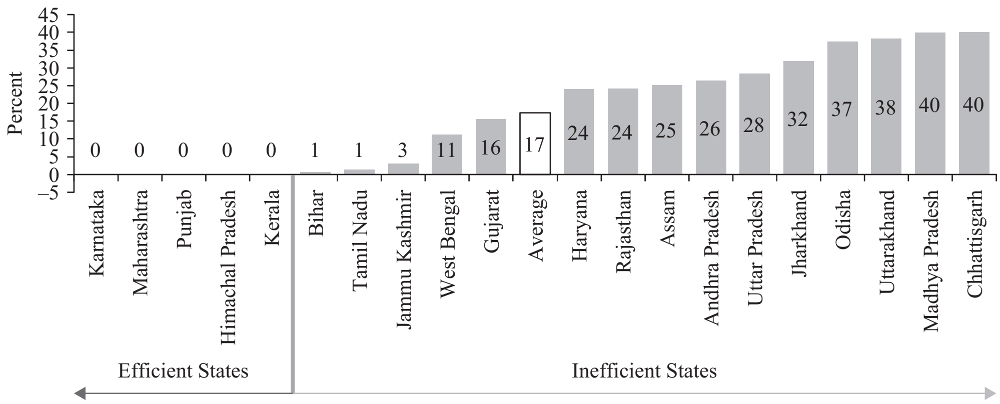
4.1.5 Information and broadcasting

We have used one input and two outputs for analyzing the efficiency of information and broadcasting. The input is the public expenditure on information and broadcasting as a percentage of GSDP. Due to data constraints, we have used 'Telephones per 100 population' and 'GSVA by the service sector as a percentage of GSDP' as two outputs for the output-oriented DEA analysis. It is to be noted that the data on telephones per 100 population for individual north-eastern States (except Assam) are not available. Therefore, we have excluded the other north-eastern States while analyzing the efficiency. Thus, 20 States are included in this analysis, and the results are given in Table 17.5 and Figure 17.5.[16]

Table 17.5 Information and broadcasting efficiency of Indian States (output oriented)

	Output Oriented_2005		Output Oriented_2015		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Andhra Pradesh	0.83	8	0.74	14	Himachal Pradesh, Kerala.
Assam	0.78	11	0.75	13	Karnataka, Kerala.
Bihar	0.93	5	0.99	6	Karnataka.
Chhattisgarh	0.59	20	0.60	20	Karnataka, Kerala.
Gujarat	0.71	17	0.84	10	Himachal Pradesh, Punjab.
Haryana	0.78	12	0.76	11	Himachal Pradesh, Kerala.
Himachal Pradesh	0.71	18	1.00	1	Himachal Pradesh.
Jammu Kashmir	0.74	15	0.97	8	Karnataka.
Jharkhand	0.61	19	0.68	16	Karnataka.
Karnataka	0.85	7	1.00	1	Karnataka.
Kerala	1.00	1	1.00	1	Kerala.
Madhya Pradesh	0.76	14	0.60	19	Karnataka, Kerala.
Maharashtra	1.00	1	1.00	1	Maharashtra.
Odisha	0.73	16	0.63	17	Karnataka, Kerala.
Punjab	1.00	1	1.00	1	Punjab.
Rajasthan	0.76	13	0.76	12	Karnataka, Kerala.
Tamil Nadu	0.95	4	0.99	7	Kerala, Maharashtra, Punjab.
Uttar Pradesh	0.79	10	0.72	15	Karnataka.
Uttarakhand	0.83	9	0.62	18	Karnataka, Kerala.
West Bengal	0.92	6	0.89	9	Karnataka, Kerala.
Average	0.81		0.83		

Source: Authors' calculation.



Source: Authors' calculation.

Figure 17.5 Information and broadcasting inefficiency, 2018

In 2005, three states (Kerala, Maharashtra and Punjab) were efficient, while in 2018, five States (with the addition of Himachal Pradesh and Karnataka) were efficient. The four bottom States are Odisha, Uttarakhand, Madhya Pradesh and Chhattisgarh. Chhattisgarh has remained the worst performing State over 2005–18. The output-oriented efficiency score of the selected States (on average) is 0.83, which means that they could add 17 percent to their current level of output using the current input. Ten States are performing worse than the average of all States. Figure 17.5 displays the inefficiency of the selected States (Bihar and Tamilnadu-1 percent, West Bengal-11 percent, Gujarat-16 percent, Haryana and Rajasthan-24 percent, Assam-25 percent, Andhra Pradesh-26 percent, Uttar Pradesh-28 percent, Jharkhand-32 percent, Odisha-37 percent, Uttarakhand-38 percent, Madhya Pradesh and Chhatishgarh-40 percent). However, overall, the information and broadcasting efficiency of all States has improved marginally (0.81 to 0.83) from 2005 to 2018.

4.2 Economic Services

Like social sector services, budgetary allocations for economic services are essential for driving economic growth, job creation and facilitating a higher standard of living in an economy. Thus, an attempt is made here to analyze the efficiency of economic services. In this section, the public sector spending efficiencies of rural development, irrigation and flood control, energy and road transport are estimated.

4.2.1 Rural development

Agriculture and rural roads play an important role in rural development. We have chosen 'GSVA by the Agriculture sector as a percentage of GSDP' and the length of rural roads[17] to total road length as two outputs, and the expenditure on rural development as a percentage of GSDP as an input in the output-oriented DEA analysis. The information on rural roads is available up to 2016 (latest). Therefore, instead of 2018, we have calculated the efficiency for 2016 using the latest information. The results are shown in Table 17.6 and Figure 17.6.

The selected DEA approach finds that Assam, Goa, Punjab, Madhya Pradesh and Tripura were the most efficient in rural development spending in 2016. Uttarakhand was the least efficient State in this sector. Many states, such as Arunachal Pradesh, Assam, Bihar, Chhattisgarh, Goa, Gujarat, Himachal Pradesh, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Meghalaya, Mizoram, Nagaland, Odisha, Rajasthan, Tamil Nadu, Tripura, Uttarakhand and West Bengal, have improved their efficiency during 2005–16. Overall, on average, the output efficiency score has also drastically improved from 0.64 to 0.76 during this period. However, still, there is scope for improvement using the existing resources (Figure 17.6). For example, Tamil Nadu is producing 28 percent less output with the existing resources. The inefficiency chart (Figure 17.6) displays significant scope for improvement in the rural development output for many States.

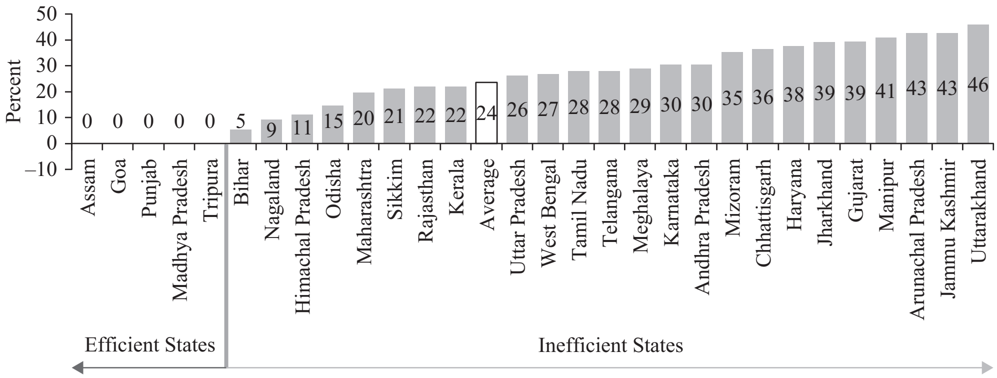
4.2.2 Irrigation and flood control

The efficiency of irrigation and flood control is calculated by using two outputs, the 'ratio of gross irrigated area to the gross shown area' and the 'yield of Rabi food grains', and one input of public expenditure, 'irrigation and flood control' as a percentage of GSDP. The latest information related to the gross irrigated area is available for the year 2017. That is why we have estimated the efficiency of irrigation and flood control for 2005 and 2017. The results are displayed in Table 17.7 and Figure 17.7.

Table 17.6 Rural development efficiency of Indian States (output oriented)

	Output Oriented_2005		Output Oriented_2016		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Andhra Pradesh	0.74	10	0.70	20	Assam, Madhya Pradesh, Tripura.
Arunachal Pradesh	0.54	20	0.57	27	Assam, Madhya Pradesh, Tripura.
Assam	0.79	8	1.00	1	Assam.
Bihar	0.83	7	0.95	6	Tripura.
Chhattisgarh	0.59	18	0.64	22	Assam, Madhya Pradesh, Tripura.
Goa	0.20	28	1.00	1	Goa.
Gujarat	0.48	22	0.61	25	Assam, Madhya Pradesh, Punjab.
Haryana	0.76	9	0.62	23	Assam, Madhya Pradesh, Punjab.
Himachal Pradesh	0.86	5	0.89	8	Assam, Tripura.
Jammu Kashmir	0.71	12	0.57	28	Assam, Madhya Pradesh, Tripura.
Jharkhand	0.48	23	0.61	24	Tripura.
Karnataka	0.56	19	0.70	19	Assam, Goa, Punjab.
Kerala	0.46	25	0.78	13	Assam, Goa.
Madhya Pradesh	0.84	6	1.00	1	Madhya Pradesh.
Maharashtra	0.28	27	0.80	10	Assam, Goa, Punjab.
Manipur	0.63	15	0.59	26	Tripura.
Meghalaya	0.59	17	0.71	18	Tripura.
Mizoram	0.53	21	0.65	21	Assam, Tripura.
Nagaland	0.87	4	0.91	7	Assam, Madhya Pradesh, Tripura.
Odisha	0.62	16	0.85	9	Assam, Tripura.
Punjab	1.00	1	1.00	1	Punjab.
Rajasthan	0.73	11	0.78	12	Assam, Madhya Pradesh, Tripura.
Sikkim	1.00	1	0.79	11	Assam, Tripura.
Tamil Nadu	0.33	26	0.72	16	Assam, Goa.
Telangana	-	-	0.72	17	Assam, Tripura.
Tripura	0.67	13	1.00	1	Tripura.
Uttar Pradesh	0.89	3	0.74	14	Assam, Madhya Pradesh, Tripura.
Uttarakhand	0.47	24	0.54	29	Assam, Tripura.
West Bengal	0.64	14	0.73	15	Assam, Madhya Pradesh, Tripura.
Average	**0.64**		**0.76**		

Source: Authors' calculation.



Source: Authors' calculation.

Figure 17.6 Rural development inefficiency, 2016

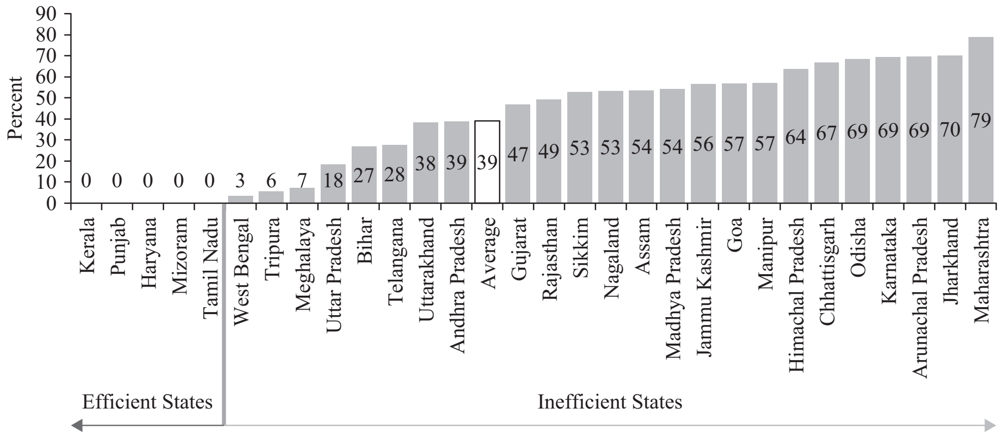
Table 17.7 Irrigation and flood control efficiency of Indian States (output oriented)

	Output Oriented_2005		Output Oriented_2017		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Andhra Pradesh	0.64	10	0.61	13	Punjab.
Arunachal Pradesh	0.33	23	0.31	27	Punjab.
Assam	0.41	19	0.46	18	Punjab.
Bihar	0.60	11	0.73	10	Punjab.
Chhattisgarh	0.25	27	0.33	24	Punjab.
Goa	0.26	26	0.43	21	Punjab.
Gujarat	0.55	13	0.53	14	Punjab.
Haryana	0.88	5	1.00	1	Haryana.
Himachal Pradesh	0.48	16	0.36	23	Punjab.
Jammu Kashmir	0.43	18	0.44	20	Punjab.
Jharkhand	0.28	25	0.30	28	Punjab.
Karnataka	0.35	20	0.31	26	Punjab.
Kerala	0.87	6	1.00	1	Kerala.
Madhya Pradesh	0.31	24	0.46	19	Punjab.
Maharashtra	0.19	28	0.21	29	Punjab.
Manipur	0.79	7	0.43	22	Punjab.
Meghalaya	1.00	1	0.93	8	Kerala, Punjab.
Mizoram	0.53	14	1.00	1	Mizoram.
Nagaland	0.56	12	0.47	17	Haryana, Kerala, Tamil Nadu.
Odisha	0.34	21	0.31	25	Punjab.
Punjab	1.00	1	1.00	1	Punjab.

Table 17.7 (Continued)

	Output Oriented_2005		Output Oriented_2017		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Rajasthan	0.48	15	0.51	15	Punjab.
Sikkim	0.34	22	0.47	16	Kerala, Mizoram, Tamil Nadu.
Tamil Nadu	0.93	4	1.00	1	Tamil Nadu.
Telangana			0.72	11	Punjab.
Tripura	0.69	9	0.94	7	Haryana, Kerala, Tamil Nadu.
Uttar Pradesh	0.77	8	0.82	9	Punjab.
Uttarakhand	0.46	17	0.62	12	Haryana, Kerala, Punjab.
West Bengal	1.00	1	0.97	6	Haryana, Kerala, Tamil Nadu.
Average	**0.56**		**0.61**		

Source: Authors' calculation.



Source: Authors' calculation.

Figure 17.7 Irrigation and flood control inefficiency, 2017

The result shows that, in the year 2017, the most efficient States were Kerala, Punjab, Haryana, Mizoram and Tamil Nadu (five States). Punjab was the only State which performed very efficiently in both 2005 and 2017. Although there is an improvement in the overall States' performance from 2005 to 2017 (from 0.56 to 0.61), many states are using their resources very inefficiently, and output could be enhanced by nearly 40 percent with the current level of spending. States like Odisha, Karnataka, Arunachal Pradesh, Jharkhand and Maharashtra are the worst performers. These States are producing nearly 30 percent or less of optimal output with their current level of spending. Almost 55 percent of States (16 States) are producing irrigation and flood control output far below the all-State average. The inefficiency level ranges between 53 and 79 percent for 14 States (Figure 17.7). The efficiency level of Assam, Bihar, Chhattisgarh, Goa, Haryana, Jammu Kashmir, Jharkhand, Kerala, Madhya Pradesh, Maharashtra, Mizoram, Rajasthan, Sikkim, Tamil Nadu, Tripura, Uttar Pradesh and Uttarakhand improved from 2005 to 2017. The efficiency of the remaining States (except Punjab) has deteriorated during that period. Thus, many States' output could be improved by following their peer groups and adopting best practices.

4.2.3 Energy

For measuring energy efficiency, two outputs are selected, that is, per capita availability of power (Kilowatt-Hour) and installed capacity of power per thousand population (megawatt). Similarly, for input, we have chosen the public budgetary outlay on the energy to GSDP ratio. The energy efficiency is calculated for 2005 and 2018 across 29 Indian States. The results are displayed in Table 17.8 and Figure 17.8.

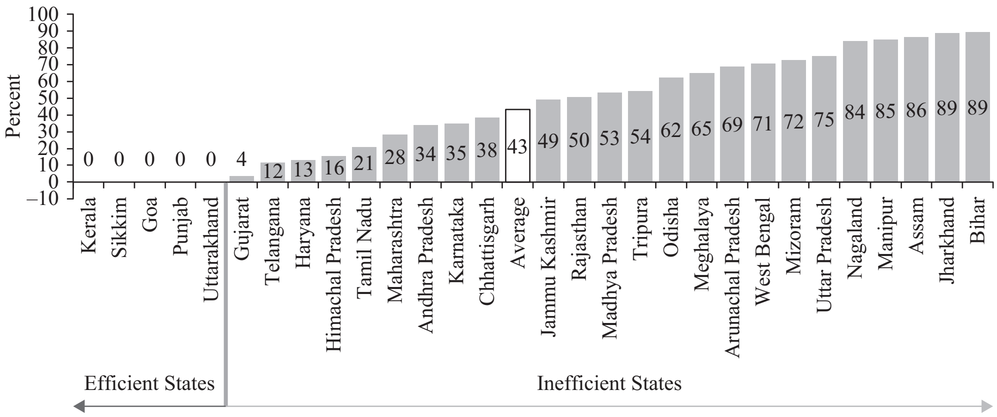
Table 17.8 Energy efficiency of Indian States (output oriented)

	Output Oriented_2005		Output Oriented_2018		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Andhra Pradesh	0.61	14	0.66	12	Punjab, Uttarakhand.
Arunachal Pradesh	0.60	15	0.31	21	Goa, Sikkim.
Assam	0.16	27	0.14	27	Goa, Punjab.
Bihar	0.15	28	0.11	29	Goa, Punjab, Sikkim.
Chhattisgarh	0.86	8	0.62	14	Goa, Punjab, Sikkim.
Goa	1.00	1	1.00	1	Goa.
Gujarat	0.86	9	0.96	6	Goa, Punjab, Sikkim.
Haryana	0.84	10	0.87	8	Goa, Punjab, Sikkim.
Himachal Pradesh	1.00	1	0.84	9	Punjab, Sikkim, Uttarakhand.
Jammu Kashmir	0.56	16	0.51	15	Goa, Sikkim.
Jharkhand	0.27	24	0.11	28	Goa, Punjab, Sikkim.
Karnataka	0.65	12	0.65	13	Goa, Punjab, Sikkim.
Kerala	1.00	1	1.00	1	Kerala.
Madhya Pradesh	0.39	21	0.47	17	Goa, Punjab, Sikkim.
Maharashtra	0.93	6	0.72	11	Goa, Punjab, Sikkim.
Manipur	0.27	23	0.15	26	Goa, Punjab, Sikkim.
Meghalaya	0.48	19	0.35	20	Goa, Punjab, Sikkim.
Mizoram	0.48	18	0.28	23	Goa, Sikkim.
Nagaland	0.19	26	0.16	25	Goa, Punjab, Sikkim.
Odisha	0.92	7	0.38	19	Goa, Punjab, Sikkim.
Punjab	1.00	1	1.00	1	Punjab.
Rajasthan	0.51	17	0.50	16	Goa, Sikkim.
Sikkim	0.77	11	1.00	1	Sikkim.
Tamil Nadu	1.00	1	0.79	10	Goa, Punjab, Sikkim.

Table 17.8 (Continued)

	Output Oriented_2005		Output Oriented_2018		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Telangana	-	-	0.88	7	Goa, Punjab, Sikkim.
Tripura	0.28	22	0.46	18	Sikkim, Uttarakhand.
Uttar Pradesh	0.27	25	0.25	24	Goa, Punjab, Sikkim.
Uttarakhand	0.64	13	1.00	1	Uttarakhand.
West Bengal	0.43	20	0.29	22	Punjab, Uttarakhand.
Average	**0.61**		**0.57**		

Source: Authors' calculation.



Source: Authors' calculation.

Figure 17.8 Energy inefficiency, 2018

It is seen that only five states, Kerala, Sikkim, Goa, Punjab and Uttarakhand, were energy efficient States in 2018. Jharkhand and Bihar were very poor performing States in terms of energy output. Punjab, Kerala and Goa were consistently performing better in both 2005 and 2018. On average, States are producing only 57 percent of their potential output, implying another 43 percent of output can be achieved without increasing the current energy spending. Nearly 30 percent of States are producing less than 30 percent of their potential output. Those States are operating below the all-State average, their energy performances are very low (inefficiency ranges from 50 to 90 percent). Seven States, Andhra Pradesh, Gujarat, Haryana, Madhya Pradesh, Sikkim, Tripura and Uttarakhand, have improved their energy efficiency during these selected years. Thus, there appears to be large scope for most State governments to take necessary steps to reduce their inefficiencies in the energy sector.

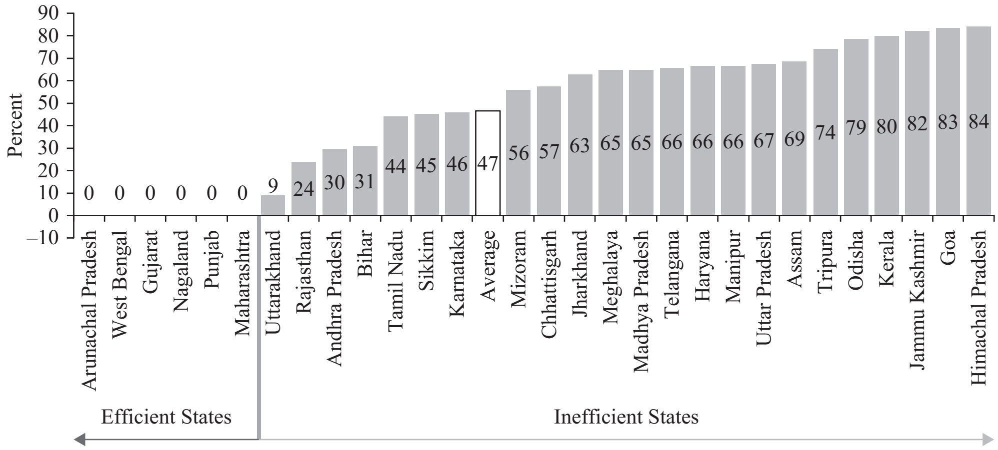
4.2.4 Road transport

Road transportation is the major form of physical infrastructure in India. For computing the efficiency of road transport, two outputs are considered, that is, the ratio of the total length of state highways to total road length and the total number of road accidents[18] per thousand population. The input is public spending on road transport to GSDP ratio.[19] As mentioned earlier, the latest information on roads is available up to 2016; the efficiency analysis is conducted for 2005 and 2016. The output-oriented DEA results are shown in Table 17.9 (efficient) and Figure 17.9 (inefficiency).

Table 17.9 Road transport efficiency of Indian States (output oriented)

	Output Oriented_2005		Output Oriented_2016		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Andhra Pradesh	0.26	24	0.70	9	Gujarat, Nagaland, West Bengal.
Arunachal Pradesh	0.21	25	1.00	1	Arunachal Pradesh.
Assam	0.31	21	0.31	23	Gujarat, Nagaland, West Bengal.
Bihar	1.00	1	0.69	10	Gujarat, Nagaland, West Bengal.
Chhattisgarh	0.34	19	0.43	15	Arunachal Pradesh, Gujarat, Nagaland.
Goa	0.18	26	0.17	28	Arunachal Pradesh, Gujarat.
Gujarat	0.87	4	1.00	1	Gujarat.
Haryana	0.63	8	0.34	20	Gujarat, Nagaland, West Bengal.
Himachal Pradesh	0.65	6	0.16	29	Arunachal Pradesh, Gujarat, Nagaland.
Jammu Kashmir	0.08	28	0.18	27	Nagaland, West Bengal.
Jharkhand	0.85	5	0.37	16	Gujarat, Nagaland, West Bengal.
Karnataka	0.55	9	0.54	13	Arunachal Pradesh, Gujarat.
Kerala	0.15	27	0.20	26	Gujarat, Nagaland, West Bengal.
Madhya Pradesh	0.39	17	0.35	18	Arunachal Pradesh, Gujarat, Nagaland.
Maharashtra	1.00	1	1.00	1	Maharashtra.
Manipur	0.55	10	0.34	21	Arunachal Pradesh, Gujarat, Nagaland.
Meghalaya	0.92	3	0.35	17	Arunachal Pradesh, Gujarat, Nagaland.
Mizoram	0.50	14	0.44	14	Arunachal Pradesh, Gujarat, Nagaland.
Nagaland	0.43	15	1.00	1	Nagaland.
Odisha	0.28	22	0.21	25	Arunachal Pradesh, Gujarat, Nagaland.
Punjab	0.39	18	1.00	1	Punjab.
Rajasthan	0.55	11	0.76	8	Gujarat, Nagaland, West Bengal.
Sikkim	0.64	7	0.55	12	Arunachal Pradesh, Gujarat, Nagaland.
Tamil Nadu	0.28	23	0.56	11	Gujarat, Maharashtra.
Telangana			0.34	19	Gujarat, Nagaland, West Bengal.
Tripura	0.31	20	0.26	24	Gujarat, Nagaland, West Bengal.
Uttar Pradesh	0.54	12	0.33	22	Gujarat, Nagaland, West Bengal.
Uttarakhand	0.50	13	0.91	7	Gujarat, Nagaland, West Bengal.
West Bengal	0.42	16	1.00	1	West Bengal.
Average	**0.49**		**0.53**		

Source: Authors' calculation.



Source: Authors' calculation.

Figure 17.9 Road transport inefficiency, 2016

As shown in Table 17.9, Gujarat, West Bengal, Arunachal Pradesh, Punjab, Nagaland and Maharashtra (six States) were efficient in 2016, while Odisha, Kerala, Jammu Kashmir, Goa and Himachal Pradesh were the bottom five states in the road transport category. The average efficiency score is 0.53, implying the huge potential for improvement. States like Andhra Pradesh, Arunachal Pradesh, Nagaland, Punjab, Uttarakhand and West Bengal significantly improved their road transport efficiency between 2005 and 2016. The efficiency level of many States, such as Bihar, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, Manipur, Meghalaya and so on, has declined during this period. As reflected in Figure 17.9, many States are very poorly managing their public expenditure on road transport (inefficient).

4.3 General Services

In this section, the efficiencies of public works and police are examined.

4.3.1 Public works

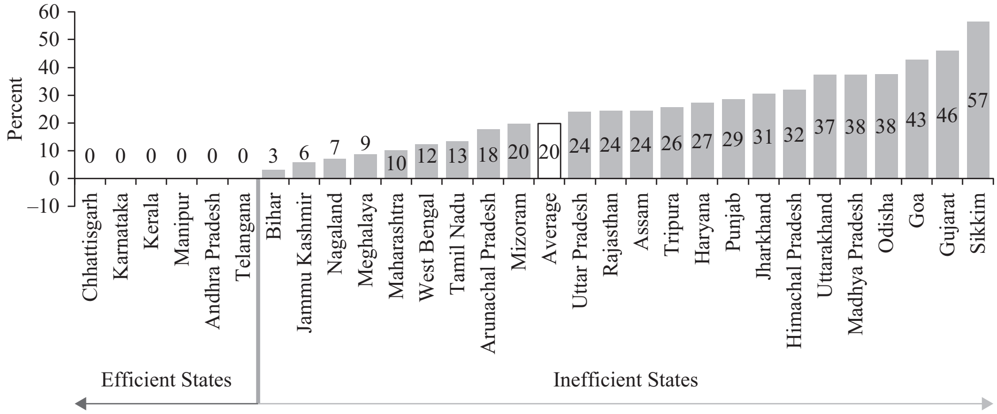
The public works department is generally involved in construction activities and overall providing public services. Due to data constraints, we have chosen two outputs, that is, GSVA by the service sector as a percentage of GSDP and GSVA by the construction sector as a percentage of GSDP, and the input is public works expenditure as a percentage of GSDP. The efficiency results of public works are shown in Table 17.10.

The results show that, in 2018, Kerala, Karnataka, Chhattisgarh, Manipur, Andhra Pradesh and Telangana were the most efficient States while Sikkim was the least efficient State. Only three states, Kerala, Manipur and Andhra Pradesh, were efficient in both 2005 and 2018. The overall efficiency in public works has deteriorated from 0.87 to 0.80. Nearly half of the States are producing less output than the all-State average. Figure 17.10 shows the public works inefficiencies of many States, which ranges between 3 and 57 percent.

Table 17.10 Public work efficiency of Indian States (output oriented)

	Output Oriented_2005		Output Oriented_2018		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Andhra Pradesh	1.00	1	1.00	1	Andhra Pradesh.
Arunachal Pradesh	0.83	18	0.82	14	Chhattisgarh, Kerala.
Assam	0.78	21	0.76	18	Kerala, Manipur.
Bihar	0.93	11	0.97	7	Kerala, Manipur.
Chhattisgarh	0.62	27	1.00	1	Chhattisgarh.
Goa	0.76	23	0.57	27	Manipur, Telangana.
Gujarat	0.78	20	0.54	28	Karnataka, Kerala, Telangana.
Haryana	0.78	22	0.73	20	Karnataka, Kerala, Manipur.
Himachal Pradesh	1.00	1	0.68	23	Kerala, Manipur.
Jammu Kashmir	0.89	14	0.94	8	Kerala, Manipur.
Jharkhand	0.61	28	0.69	22	Kerala, Manipur.
Karnataka	0.85	16	1.00	1	Karnataka.
Kerala	1.00	1	1.00	1	Kerala.
Madhya Pradesh	0.76	24	0.62	25	Chhattisgarh, Kerala.
Maharashtra	0.98	10	0.90	11	Karnataka, Kerala, Telangana.
Manipur	1.00	1	1.00	1	Manipur.
Meghalaya	0.86	15	0.91	10	Kerala, Manipur.
Mizoram	1.00	9	0.80	15	Kerala, Manipur.
Nagaland	0.91	13	0.93	9	Kerala, Manipur.
Odisha	0.76	25	0.62	26	Kerala, Manipur.
Punjab	0.71	26	0.71	21	Karnataka, Kerala, Manipur.
Rajasthan	1.00	2	0.76	17	Karnataka, Kerala, Telangana.
Sikkim	1.00	2	0.43	29	Kerala, Manipur.
Tamil Nadu	1.00	2	0.87	13	Karnataka, Kerala, Manipur.
Telangana			1.00	1	Telangana.
Tripura	1.00	2	0.74	19	Kerala, Manipur.
Uttar Pradesh	0.79	19	0.76	16	Chhattisgarh, Kerala.
Uttarakhand	0.85	17	0.63	24	Kerala, Manipur.
West Bengal	0.92	12	0.88	12	Karnataka, Kerala, Manipur.
Average	0.87		0.80		

Source: Authors' calculation.



Source: Authors' calculation.

Figure 17.10 Public work inefficiency, 2018

4.3.2 Police

Police play a crucial role in every nation. The police efficiency is measured using two outputs and one input for both 2005 and 2018. Police expenditure to GSDP is used as an input. The two outputs are 'case disposal rate' and 'charge sheeting rate' by police.[20] The case disposal rate is calculated as the ratio of total cases disposed of by police to total cases reported for investigation. Similarly, the charge sheeting rate is computed as total cases charge sheeted to total cases reported for investigation. The police efficiency results are displayed in Table 17.11.

Table 17.11 Police efficiency of Indian States (output oriented)

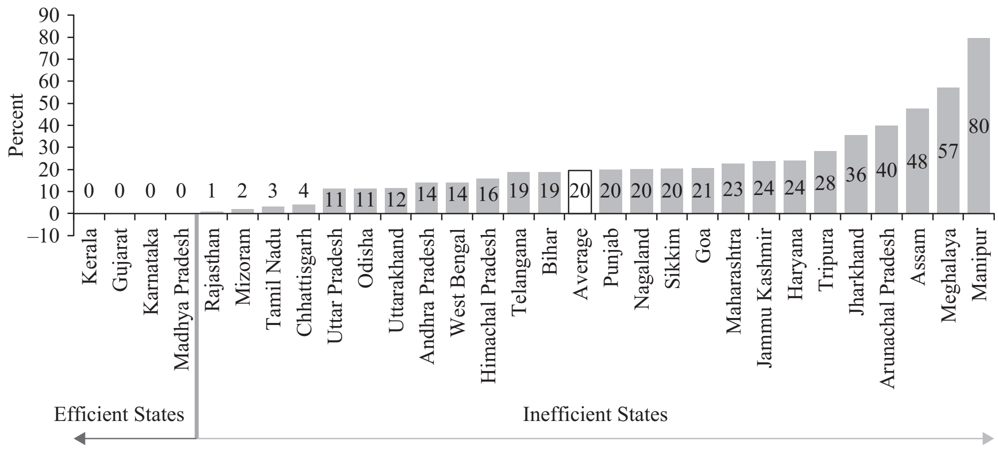
	Output Oriented_2005		Output Oriented_2018		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Andhra Pradesh	0.97	7	0.86	12	Kerala.
Arunachal Pradesh	0.75	22	0.60	26	Kerala.
Assam	0.60	26	0.52	27	Madhya Pradesh.
Bihar	0.78	20	0.81	16	Kerala.
Chhattisgarh	1.00	1	0.96	8	Gujarat, Madhya Pradesh.
Goa	0.92	10	0.79	20	Gujarat, Madhya Pradesh.
Gujarat	1.00	1	1.00	1	Gujarat.
Haryana	0.89	13	0.76	23	Gujarat, Madhya Pradesh.
Himachal Pradesh	0.90	12	0.84	14	Gujarat, Kerala.
Jammu Kashmir	0.74	23	0.76	22	Kerala.
Jharkhand	0.70	25	0.64	25	Kerala.
Karnataka	0.83	17	1.00	1	Karnataka.

(*Continued*)

Table 17.11 (Continued)

	Output Oriented_2005		Output Oriented_2018		
States	Eff. Score	Rank	Eff. Score	Rank	Peer Group
Kerala	0.91	11	1.00	1	Kerala.
Madhya Pradesh	1.00	4	1.00	1	Madhya Pradesh.
Maharashtra	0.89	14	0.77	21	Gujarat, Kerala.
Manipur	0.28	28	0.20	29	Madhya Pradesh.
Meghalaya	0.52	27	0.43	28	Kerala.
Mizoram	0.94	8	0.98	6	Gujarat, Madhya Pradesh.
Nagaland	0.74	24	0.80	18	Madhya Pradesh.
Odisha	0.93	9	0.89	10	Kerala.
Punjab	0.87	15	0.80	17	Kerala.
Rajasthan	0.99	6	0.99	5	Gujarat, Madhya Pradesh.
Sikkim	0.78	21	0.80	19	Madhya Pradesh.
Tamil Nadu	0.83	16	0.97	7	Gujarat, Kerala.
Telangana			0.81	15	Kerala.
Tripura	0.79	19	0.72	24	Kerala.
Uttar Pradesh	1.00	5	0.89	9	Madhya Pradesh.
Uttarakhand	1.00	1	0.88	11	Kerala.
West Bengal	0.80	18	0.86	13	Kerala.
Average	0.83		0.80		

Source: Authors' calculation.



Source: Authors' calculation.

Figure 17.11 Police inefficiency, 2018

The result shows that, out of 29 States, only four States were efficient in 2018. These efficient States are Kerala, Gujarat, Karnataka and Madhya Pradesh. Four north-eastern States, that is, Assam, Arunachal Pradesh, Meghalaya and Manipur, were the least efficient States. Madhya Pradesh and Gujarat are the only States which were efficient in both 2005 and 2018. The average police efficiency score of all States is 0.80, implying that output could be enhanced by 20 percent using the current public spending. Some of these States, like Bihar, Jammu Kashmir, Karnataka, Kerala, Mizoram, Nagaland, Sikkim, Tamil Nadu and West Bengal, have marginally improved during 2005 and 2018. The inefficiency level of Telangana, Bihar, Punjab, Nagaland, Sikkim and Goa is almost similar to the all-State average (Figure 17.11). The given figure also shows that the majority of States are inefficient. The inefficient States might follow their respective peer groups to enhance their output with the current level of spending.

A summary of the selected public sector efficiency rankings is given in Table 17.12. Out of 11 public sectors, the efficiency trend has improved for water supply and sanitation, information and broadcasting, rural development, irrigation and flood control and road transport during 2005 and 2018. Public spending on water supply and sanitation is used most efficiently whereas road transport is the least efficient.

5. AGGREGATE PUBLIC SECTOR PERFORMANCE OF ALL STATES

After analyzing separately the efficiency of the selected public sector spendings, we are interested in evaluating the respective States' overall performances. The overall public sector performance index[21] for 2018 is computed using the Slacks-Based Measure of efficiency (SBM) model developed by Tone (2001). However, to obtain the aggregate public sector performance index, first, we compute the overall social sector efficiency index,[22] and combined economic and general sector efficiency index[23] using the SBM approach. The empirical results with ranks are shown in Table 17.13.

In the social sector index, two western States (Goa and Maharashtra), two northern States (Himachal Pradesh and Punjab) and one north-eastern State (Nagaland) are efficient. However, the worst-performing States in the social sector are two each from the eastern region (Bihar and Jharkhand) and the central region (Madhya Pradesh and Uttar Pradesh). The remaining eastern and central region States are also performing below the all-State average. Rajasthan from the north and Andhra Pradesh from the south are poor performing States compared to other States of their region in the social sector.

Similarly, the results of the combined economic and general sector performance index show that four southern States (Andhra Pradesh, Karnataka, Kerala and Tamil Nadu), two each from western States (Gujarat and Maharashtra) and north-eastern States (Mizoram and Nagaland), one each from northern (Punjab), eastern (West Bengal) and central (Uttarakhand) States are efficient. Compared to the central States, eastern States have fared poorly in this sector. Most of the southern States and western States have better performance than the other regions in the economic and general sector.

Further, the aggregate performance index shows that Kerala, Maharashtra and Punjab are the most efficient in using their public spending, while eastern states such as Jharkhand, Bihar and Odisha along with two north-eastern States (Assam and Manipur) are very inefficient in their public spending. Except for the most efficient States (Kerala, Maharashtra and Punjab), the other three better-performing States are Gujarat, Tamil Nadu and Karnataka. Among the

Table 17.12 Summary of DEA results (output efficiency scores)

Sector		2005	2018 or latest	Sector	2005	2018 or latest
Education	Number	7	7	**Irrigation and Flood Control**	3	5
	Efficient States	Haryana, Himachal Pradesh, Kerala, Maharashtra, Meghalaya, Tamil Nadu, and Uttarakhand	Haryana, Himachal Pradesh, Kerala, Punjab, Sikkim, Tamil Nadu, and Telangana		Punjab, West Bengal, and Meghalaya	Haryana, Kerala, Mizoram, Punjab, and Tamil Nadu
	Average	0.88	0.79		0.56	0.61
	Median	0.90	0.83		0.50	0.51
	Maximum	1	1		1	1
	Minimum	0.61	0.48		0.19	0.21
Health	Number	8	6	**Energy**	5	5
	Efficient States	Haryana, Himachal Pradesh, Kerala, Chhattisgarh, Gujarat, Maharashtra, Manipur, and Mizoram	Himachal Pradesh, Karnataka, Kerala, Nagaland, Maharashtra, Mizoram		Goa, Himachal Pradesh, Kerala, Punjab, and Tamil Nadu	Goa, Kerala, Punjab, Sikkim, and Uttarakhand
	Average	0.77	0.74		0.61	0.57
	Median	0.78	0.80		0.60	0.51
	Maximum	1	1		1	1
	Minimum	0.42	0.35		0.15	0.11

Table 17.12 (Continued)

Sector		2005	2018 or latest	Sector	2005	2018 or latest
Water Supply and Sanitation	Number	5	8	**Road Transport**	2	6
	Efficient States	Kerala, Punjab, Mizoram, Tripura, and West Bengal	Kerala, Maharashtra, Goa, Punjab, Sikkim, Manipur, Nagaland, and Mizoram		Bihar and Maharashtra	Arunachal Pradesh, Gujarat, Maharashtra, Nagaland, Punjab, and West Bengal
	Average	0.85	0.90		0.49	0.53
	Median	0.88	0.97		0.46	0.43
	Maximum	1	1		1	1
	Minimum	0.44	0.55		0.08	0.16
Social Welfare and Nutrition	Number	5	4	**Public Works**	8	6
	Efficient States	Kerala, Goa, Punjab, Manipur, and Mizoram	Kerala, Gujarat, Punjab, and Mizoram		Andhra Pradesh, Himachal Pradesh, Kerala, Manipur, Rajasthan, Sikkim, Tamil Nadu, and Tripura	Andhra Pradesh, Chhattisgarh, Karnataka, Kerala, Manipur, and Telangana
	Average	0.72	0.71		0.87	0.80
	Median	0.68	0.67		0.88	0.80
	Maximum	1	1		1	1
	Minimum	0.46	0.45		0.61	0.43

(*Continued*)

Table 17.12 (Continued)

Sector		2005	2018 or latest	Sector	2005	2018 or latest
	Number	3	5		3	4
Information and Broadcasting	Efficient States	Kerala, Maharashtra, and Punjab	Karnataka, Kerala, Himachal Pradesh, Maharashtra, and Punjab	**Police**	Chhattisgarh, Gujarat, and Uttarakhand	Karnataka, Kerala, Gujarat, and Madhya Pradesh
	Average	0.81	0.83		0.83	0.80
	Median	0.79	0.80		0.88	0.81
	Maximum	1	1		1	1
	Minimum	0.59	0.60		0.28	0.20
	Number	2	5			
Rural Development	Efficient States	Punjab and Sikkim	Goa, Punjab, Assam, Madhya Pradesh, and Tripura			
	Average	0.64	0.76			
	Median	0.63	0.73			
	Maximum	1	1			
	Minimum	0.20	0.54			

Source: Authors' calculation.

Table 17.13 Combined public sector efficiency of Indian States

	Social Sector		Economic and General Sector		Aggregate Public Sector	
States	Efficiency	Rank	Efficiency	Rank	Efficiency	Rank
Andhra Pradesh	0.70	21	1.00	1	0.82	12
Arunachal Pradesh	0.78	17	0.60	18	0.68	17
Assam	0.68	24	0.34	27	0.45	28
Bihar	0.56	28	0.49	23	0.52	26
Chhattisgarh	0.71	19	0.77	14	0.74	16
Goa	1.00	1	0.43	24	0.60	22
Gujarat	0.88	9	1.00	1	0.94	6
Haryana	0.80	13	0.70	15	0.75	15
Himachal Pradesh	1.00	1	0.42	26	0.59	23
Jammu Kashmir	0.76	18	0.52	22	0.62	21
Jharkhand	0.54	29	0.29	29	0.38	29
Karnataka	0.80	14	1.00	1	0.89	8
Kerala	1.00	1	1.00	1	1.00	1
Madhya Pradesh	0.66	26	0.64	17	0.65	19
Maharashtra	1.00	1	1.00	1	1.00	1
Manipur	0.80	15	0.34	28	0.47	27
Meghalaya	0.67	25	0.60	19	0.63	20
Mizoram	0.93	8	1.00	1	0.96	5
Nagaland	1.00	1	1.00	1	1.00	1
Odisha	0.69	23	0.42	25	0.52	25
Punjab	1.00	1	1.00	1	1.00	1
Rajasthan	0.69	22	0.86	13	0.76	14
Sikkim	0.96	7	0.66	16	0.78	13
Tamil Nadu	0.86	10	1.00	1	0.92	7
Telangana	0.83	12	0.92	12	0.87	10
Tripura	0.83	11	0.56	20	0.67	18
Uttar Pradesh	0.57	27	0.53	21	0.55	24
Uttarakhand	0.78	16	1.00	1	0.88	9
West Bengal	0.70	20	1.00	1	0.82	11
Average	**0.80**		**0.73**		**0.74**	

Source: Authors' calculation.

north-eastern States, Nagaland and Mizoram are utilizing their resources more efficiently than the other north-eastern States. Most of the central States (except Uttarakhand) are utilizing their resources below the average efficiency of all States.

What could be the possible reason for the better performance of the most efficient Indian States? Mohanty and Bhanumurthy (2020) pointed out that governance and economic growth play an important role in explaining efficiency in the education and social sectors. After checking the governance and growth for 2018, we also find that the most efficient States, Kerala (rank 1), Tamil Nadu (rank 2), Karnataka (4), Gujarat (5), Maharashtra (6), Punjab (7) and so on are the top ranked in governance as per the Public Affairs Index score 2018. Their economic growths are also in double digits, that is, Gujarat (13.1 percent), Karnataka (13.8 percent), Kerala (11.4 percent), Punjab (11.8 percent), Tamil Nadu (11.3 percent) and Maharashtra (10.5 percent). We have also found from our DEA results that all of these States have relatively better education, health and physical infrastructure. The combination of both social and economic infrastructure among these States might lead to better public spending efficiency. Political stability or strong pro-government sentiment might be another factor for improved efficiency in these States, although this could not be covered in this chapter due to data deficiency.

6. SUMMARY AND CONCLUSIONS

As the level of public expenditure, efficiency becomes a crucial factor in any budget-making process, especially when there are hard budget constraints. In this chapter, an attempt has been made to understand the extent of efficiency across public sector expenditures at the state level. Therefore, the study attempts to estimate the efficiency of 11 major public sector expenditures among 29 states: education, health, water supply and sanitation, information and broadcasting, social welfare and nutrition, rural development, irrigation and flood control, energy, roads transport, police and public works. The study uses the output-oriented DEA approach (most appropriate in the Indian context), instead of input oriented, for measuring the efficiencies of the selected public sectors. Because the current public sector spending as a percent of GSDP is very low, further reduction in public spending might not be an option. Analyzing the public sector efficiencies and their rankings will help policymakers identify the best and worst-performing states in the respective public sectors. The analysis will portray the leaders and laggards of each of these major public sectors. This might help them to fine tune policies for the development of various public sectors.

Based on the output-oriented DEA results, the states that are performing better are Haryana, Himachal Pradesh, Kerala, Punjab, Sikkim, Tamil Nadu, and Telangana in Education; Himachal Pradesh, Karnataka, Kerala, Nagaland, Maharashtra and Mizoram in Health; Kerala, Maharashtra, Goa, Punjab, Sikkim, Manipur, Nagaland and Mizoram in Water Supply and Sanitation; Kerala, Gujarat, Punjab and Mizoram in Social Welfare and Nutrition; Karnataka, Kerala, Himachal Pradesh, Maharashtra and Punjab in Information and Broadcasting; Goa, Punjab, Assam, Madhya Pradesh and Tripura in Rural Development; Haryana, Kerala, Mizoram, Punjab and Tamil Nadu in Irrigation and Flood Control; Goa, Kerala, Punjab, Sikkim and Uttarakhand in Energy; Arunachal Pradesh, Gujarat, Maharashtra, Nagaland, Punjab and West Bengal in Road Transport; Andhra Pradesh, Chhattisgarh, Karnataka, Kerala, Manipur and Telangana in Public Works, and Karnataka, Kerala, Gujarat and Madhya Pradesh in Police. The overall public sector performance shows that Kerala, Maharashtra, Nagaland and Punjab (most efficient) followed by Mizoram, Gujarat, Tamil Nadu and Karnataka (nearly 90 percent or above) are efficiently using their public expenditure, while eastern states such as Jharkhand, Bihar and Odisha along with two north-eastern States (Assam and Manipur) are most inefficient in their public spending. Most of the central States (except Uttarakhand) are utilizing their resources below the average efficiency of all States.

Our results reveal a significant variation in terms of inefficiencies across the sectors and the states as well as over the years. These inefficiencies imply that a higher budgetary sectoral allocation might not be sufficient for the production of the potential output. Public sector spending efficiency could be improved if the less efficient States would emulate the better-performing ones (peer groups). The inefficiencies of public sector spending suggest that there is a large scope for reforms that could impact the sector in such a way that those sectors/states could become efficient.

One of the major limitations of this study was the data constraints that limited the choice of sectors and variables (outputs) as well as limited the analysis in understanding what drives the divergences in the efficiency outcomes in those sectors/states. As discussed earlier, while the level of governance plays a crucial role in improving the efficiencies, due to limited information, mapping the same to sectoral outputs could not be undertaken in this chapter and this could be attempted in the future.

NOTES

1. Some more studies are reviewed in the literature review section.
2. A report on State Finances by the Reserve Bank of India (2018) suggests that "Given the funding constraints on states' budgets and rising borrowing costs, improving the efficiency of public expenditures holds the key to achieving the Fiscal Responsibility and Budget Management targets" (p. 2).
3. The average public spending as a percentage of GSDP also ranges from 11 to 64 during the last three years, that is, from 2016−17 to 2018−19. Even the individual sectoral spending has a wide variation across States.
4. However, there are some studies relating to macroeconomic and health system performance of Indian States using data envelopment analysis (see, for example, Sahoo & Acharya, 2012; Acharya & Sahoo, 2017; Acharya et al., 2019; Mohanty et al., 2021; among others).
5. Mohanty and Bhanumurthy (2020) attempted to evaluate the public spending efficiency of education, health and the overall social sector.
6. These are Andhra Pradesh, Arunachal Pradesh, Assam, Bihar, Chhattisgarh, Goa, Gujarat, Haryana, Himachal Pradesh, Jammu and Kashmir, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Manipur, Meghalaya, Mizoram, Nagaland, Odisha, Punjab, Rajasthan, Sikkim, Tamil Nadu, Telangana, Tripura, Uttar Pradesh, Uttarakhand and West Bengal. Union Territories and Islands are not included due to low reporting and data availability.
7. The reasons are as follows. The three new States, namely, Chhattisgarh, Jharkhand and Uttarakhand were formed in November 2000, and would require some initial years to function smoothly. Any prior period analysis would force us to exclude these States from our analysis. Then, India has implemented the crucial institutional reforms of Fiscal Responsibility and Budget Management (FRBM) Acts, 2003 in 2004, which has a greater implication for the decision of State financing, especially the public expenditure part. Some of our output variables are only available from 2005 onwards. Therefore, due to data constraints, we have chosen 2005 as the beginning year and 2018 or the latest information available as the end period of our analysis.
8. For a given year, we have considered two-year averages of public spending and their current year output (the study period) for the empirical analysis as expenditures and outputs are not synchronously related.
9. For example, police and public works are included under general services; education, health, water supply and sanitation, information and broadcasting and social welfare and nutrition are comprised under social services; and rural development, irrigation and flood control, energy and roads transport are encompassed under economic services of states budget.
10. However, in few cases the DEA method may underestimate inefficiencies as it assumes that all the on-line decision-making units (DMU) are efficient wherein they might have scope for improvement. Furthermore, there could be a possibility of overestimation of efficient units when there are many inputs and/or outputs relative to a small sample of DMUs (Sahoo & Mohapatra, 2001; Dutu & Sicari, 2016).
11. It is derived by adding the total number of Sub-health Centre (SCs), Primary Health Centre (PHCs) and Community Health Center (CHCs) per 1000 population.

12. It includes the expenditure made on the budget heads of “Medical and Public Health” and “Family Welfare”.
13. It is directly interpretable as the ratio of children who survived the first year to the number of children who died; which increases with better health status.
14. As the selected output data for 2005 were not available, we have used the data of 2001 as a proxy for 2005 assuming the percentage distribution of households could have been the same for the next 4 years (not changed drastically).
15. As these are considered as negative outputs, we have taken the inverse of these data in the DEA set-up.
16. In 2018, the information on telephones per 100 population for Chhattisgarh, Jharkhand and Uttarakhand were included under Madhya Pradesh, Bihar and Uttar Pradesh, respectively. Therefore, while calculating efficiency, the information on telephone per 100 population for Chhattisgarh, Jharkhand and Uttarakhand were drawn from these respective States.
17. Rural roads consist of Panchayati Raj roads, Pradhan Mantri Gram Sadak Yojana (PMGSY) roads and roads constructed by State PWDs. Due to data constraints, rural roads for 2005 include only the PMGSY roads.
18. Road accidents are chosen as these reflect the quality of roads. For example, if roads are of bad quality, then the number of accidents would be higher. As it is a negative output, we have taken the inverse of this variable in the DEA set-up.
19. Road transport is derived by additional expenditure on the budgetary heads of “roads and bridges” and “road transport”.
20. It is to be noted that the crime data refer to crimes committed under both Indian Penal Code (IPC) and Special and Local Laws (SLL) cognizable crimes.
21. The Information and Broadcasting sector is excluded while computing aggregate public sector efficiency because it includes only 20 States.
22. It is computed using the efficiency sores of Education, Health, Water Supply and Sanitation, Social Welfare and Nutrition and Rural Development.
23. It is calculated using the efficiency sores of Irrigation and Flood Control, Energy, Road Transport, Public Works and Police.

REFERENCES

Acharya, D., & Sahoo, B. K. 2017. Dynamic Macroeconomic Performance of Indian States: Some Post-Reform Evidence. In *Regional growth and sustainable development in Asia* (pp. 181–197). Springer, Cham.

Acharya, D., Sahoo, B. K., & Venkatachalapathy, T. K. 2019. Health System and Health Expenditure Productivity Changes in Indian States: Has It Changed for the Better in the Post-reform Period? In *Disease, Human Health, and Regional Growth and Development in Asia* (pp. 71–87). Springer, Singapore.

Adam, A., Delis, M., and Kammas, P. 2011. Public sector efficiency: leveling the playing field between OECD countries. *Public Choice*, 146(1–2), 163–183.

Afonso, A., and Aubyn, M. S. 2005. Non-parametric approaches to education and health efficiency in OECD countries. *Journal of Applied Economics*, 8(2), 227–246.

Afonso, A., Fernandes, S., 2008. Assessing and explaining the relative efficiency of local government. *Journal of Socio-Economics*, 37(5), 1946–1979.

Afonso, A., Schuknecht, L., and Tanzi, V. 2005. Public sector efficiency: an international comparison. *Public Choice*, 123(3–4), 321–347.

Afonso, A., Schuknecht, L., and Tanzi, V. 2010. Public sector efficiency: evidence for new EU member states and emerging markets. *Applied Economics*, 42(17), 2147–2164.

Angelopoulos, K., Philippopoulos, A., Tsionas, E. 2008. Does public sector efficiency matter? Revisiting the relation between fiscal size and economic growth in a world sample. *Public Choice*, 137(1–2), 245–278.

Antonelli, M. A., and De Bonis, V. 2019. The efficiency of social public expenditure in European countries: a two-stage analysis. *Applied Economics*, 51(1), 47–60.

Balaguer-Coll, M.T., Prior, D., Tortosa-Ausina, E. 2010. Decentralization and efficiency of local government. *Annals of Regional Science*, 45(3), 571–601.

Borge, L. E., Falch, T., and Tovmo, P. 2008. Public sector efficiency: the roles of political and budgetary institutions, fiscal capacity, and democratic participation. *Public Choice*, 136(3–4), 475–495.

Chan, S. G., and Karim, M. Z. A. 2012. Public spending efficiency and political and economic factors: evidence from selected East Asian countries. *Economic Annals*, 57(193), 7–24.

Charnes, A., Cooper, W. W., & Rhodes, E. 1978. Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6), 429–444

Da Cruz, N. F., Marques, R. C. 2014. Revisiting the determinants of local government performance. *Omega*, 44, 91–103.

De Borger, B., Kerstens, K. 1996. Cost efficiency of Belgian local governments: a comparative analysis of FDH, DEA, and econometric approaches. *Regional Science and Urban Economics*, 26(2), 145–170.

Doumpos, M., Cohen, S. 2014. Applying Data Envelopment Analysis on accounting data to assess and optimize the efficiency of Greek local governments. *Omega*, 46, 74–85.

Dutu, R. and Sicari, P. 2016, Public spending efficiency in the OECD: benchmarking health care, education and general administration. *OECD Economics Department Working Papers No. 1278*, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jm3st732jnq-en.

Feeny, S., and Rogers, M. 2008. Public sector efficiency, foreign aid and small island developing states. *Journal of International Development*, 20(4), 526–546.

Geys, B., Heinemann, F., Kalb, A. 2010. Voter involvement, fiscal autonomy and public sector efficiency: evidence from German municipalities. *European Journal of Political Economy*, 26(2), 265–278.

Grossman, P. J., Mavros, P., Wassmer, R. W. 1999. Public sector technical inefficiency in large US cities. *Journal of Urban Economics* 46(2), 278–299.

Gupta, S., and Verhoeven, M. 2001. The efficiency of government expenditure: experiences from Africa. *Journal of Policy Modeling*, 23(4), 433–467.

Hauner, D. 2008. Explaining differences in public sector efficiency: evidence from Russia's regions. *World Development*, 36(10), 1745–1765.

Hwang, J., and Akdede, S. H. 2011. The influence of governance on public sector efficiency: a cross-country analysis. *The Social Science Journal*, 48(4), 735–738.

Joumard, I., Andre, C., and Nicq, C. 2010. Health care systems: efficiency and institutions. *OECD Economics Department Working Papers No. 769*, OECD Publishing, Paris. http://dx.doi.org/10.2139/ssrn.1616546.

Kalb, A. 2010. The impact of intergovernmental grants on cost efficiency: theory and evidence from German municipalities. *Economic Analysis and Policy*, 40(1), 23–48.

Kalb, A., Geys, B., Heinemann, F. 2012. Value for money? German local government efficiency in a comparative perspective. *Applied Economics*, 44(2), 201–218.

Kalseth, J., Rattso, J. 1995. Spending and overspending in local government administration: a minimum requirement approach applied to Norway. *European Journal of Political Economy*, 11(2), 239–251.

Lo Storto, C. 2016. The trade-off between cost efficiency and public service quality: a non-parametric frontier analysis of Italian major municipalities. *Cities*, 51, 52–63.

Loikkanen, H. A., Susiluoto, I. 2005. Cost efficiency of Finnish municipalities in basic service provision 1994-2002. *Urban Public Economics Review*, 4, 39–64.

Mohanty, R. K., and Bhanumurthy, N. R. 2020. Assessing public expenditure efficiency at the subnational level in India: does governance matter? *Journal of Public Affairs*, 21(2), e2173. https://doi.org/10.1002/pa.2173.

Mohanty, R. K., Sahoo, B. K., and Chaudhury, P. K. 2020. Assessing the (eco) macroeconomic performance index of India: a data envelopment analysis approach. *Journal of Public Affairs*, 21(1), e2122.

Narbon-Perpina, I., and De Witte, K. 2018. Local governments' efficiency: a systematic literature review-part I. *International Transactions in Operational Research*, 25(2), 431–468.

Ouertani, M. N., Naifar, N., and Haddad, H. B. 2018. Assessing government spending efficiency and explaining inefficiency scores: DEA-bootstrap analysis in the case of Saudi Arabia. *Cogent Economics and Finance*, 6(1), 1–16.

Pérez-López, G., Prior, D., and Zafra-Gómez, J. L. 2015. Rethinking new public management delivery forms and efficiency: long-term effects in Spanish local government. *Journal of Public Administration Research and Theory*, 25(4), 1157–1183.

Reserve Bank of India. 2018. State finances: a study of budgets of 2017-18 and 2018-19. Mumbai: India.

Revelli, F. 2010. Spend more, get more? an inquiry into English local government performance. *Oxford Economic Papers*, 62(1), 185–207.

Sahoo, B. K., and Acharya, D. 2012. Constructing macroeconomic performance index of Indian states using DEA. *Journal of Economic Studies*, *39*(1), 63–83.

Sahoo, B. K., and Gstach, D. 2011. Scale economies in Indian commercial banking sector: evidence from DEA and translog estimates. *International Journal of Information Systems and Social Change*, *2*(4), 13–30.

Sahoo, B. K., and Mohapatra, P. K. 2001. Measuring technical efficiency in data envelopment analysis. *Opsearch*, *38*(5), 456–483.

Sampaio de Sousa, M.C., and Stosic, B. 2005. Technical efficiency of the Brazilian municipalities: correcting nonparametric frontier measurements for outliers. *Journal of Productivity Analysis*, 24(2), 157–181.

Saxena, M., Chotia, V., and Rao, N. V. M. 2018. Estimating the efficiency of public infrastructure investment: a state-wise analysis. *Global Business Review*, 19(4), 1037–1049.

Sutherland, D., Price, R., and Gonand, F. 2010. Improving public spending efficiency in primary and secondary education. *OECD Journal: Economic Studies*, 2009(1), 1–30.

Tone, K. 2001. A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, 130(3), 498–509.

Worthington, A.C. 2000. Cost efficiency in Australian local government: a comparative analysis of mathematical programming and econometrical approaches. *Financial Accountability and Management*, 16(3), 201–223.

Worthington, A.C., Dollery, B.E. 2000. An empirical survey of frontier efficiency measurement techniques in local government. *Local Government Studies* 26(2), 23–52.

Yadava, A. K., and Neog, Y. 2019. Public sector performance and efficiency assessment of Indian states. *Global Business Review*. https://doi.org/10.1177/0972150919862664.

18. Estimating returns to scale for the science and technology activities of Project 985 universities in China

Yao-yao Song, Xian-tong Ren, Guo-liang Yang, and Zhong-cheng Guan

1. INTRODUCTION

One of the important issues in science and technology (S&T) management departments in China is the efficiency of scientific and technological resource utilisation. The national financial allocation for S&T is 1340.8 billion Yuan (RMB) (US$219.6 billion) from 2007 to 2011, which is 2.7 times that for the period 2001 to 2005. The average annual growth rate is 23 per cent. At the same time, the number of R&D staff in China was nearly 2.6 million people in 2010, ranking first in the world. However, as public resources are still scarce and limited, the efficiency of S&T resource utilisation must be further improved. Thus, how to use these resources effectively and efficiently is currently a topic of public concern and a crucial issue facing S&T management departments. It is important that these issues, which include the rational allocation of the limited S&T resources to maximise the efficiencies of resource utilisation, be understood and mastered at the national macro-S&T management levels. Universities are an important base for training high-level innovative talented people in China, and as such, they are one of the main forces promoting and conducting basic research and high-tech original innovation. They are also a vital force for solving major national S&T issues and achieving technology transfer. Therefore, to address these issues, the returns to scale (RTS) and efficiencies of S&T resource utilisation of research universities should be investigated first.

Since the reform and opening up of China three decades ago, universities have played an increasingly important role in China's S&T development and have become an important part of the national innovation system. With the deepening of China's reform of the education system and the intense emphasis on higher education, research projects and funds for scientific research have grown substantially.

From 2005 to 2009, research funding of universities in China increased from 24.23 billion Yuan (RMB) (US$3.9689 billion) to 45.62 billion Yuan (RMB) (US$7.4726 billion), an annual growth rate of more than 17 per cent. Under the central government's S&T plan, universities undertook approximately 28 000 projects and received funding totalling approximately 27.75 billion Yuan (RMB) (US$4.5455 billion), which accounted for more than a quarter (25.7 per cent) of the total funding of the national S&T programme from the central government, a figure that has continued to grow in recent years. For example, in one part of basic research plan, called the '973 plan' or the National Key Basic Research Programme, the proportion of the funding of projects undertaken by universities exceeds 50 per cent. Taking a university as an example, its research funding income approached 2.752 billion Yuan (RMB) (US$0.4508 billion) in 2010, an increase of 38.84 per cent over the previous year, and the scale of funds accounted for 44.69 per cent of the total revenue of this particular university. Accordingly, as it can be concluded that research funding has become an important source of funding for the university sector in China, the scientific and effective management of research funding to enhance the efficient use of funds for scientific research becomes increasingly important.

In May 1998, General Secretary Ze-min Jiang of China declared in a meeting celebrating the 100th anniversary of Beijing University, ‘In order to achieve modernisation, China must have a number of the world’s leading universities at the advanced level.’ To implement the strategies developed to rejuvenate the country through science and education in a response to the CPC Central Committee and Jiang’s call, the Ministry of Education of China called for the implementation of the ‘Educational Revitalisation Action Plan for the 21st Century’, which focuses on creating world-class high-level universities, that is, the ‘985’ project. The ‘985’ project includes 39 universities, including Tsinghua University, Peking University, Xiamen University, Nanjing University, Fudan University and so on. In this chapter, we denote these universities, with the exception of the National University of Defence Science and Technology, as Project 985 universities.

The Project 985 universities, as the main bearers of research activities, represent the highest level of university research in China. A considerable proportion of university research funds flows into the Project 985 universities, with annual research funding for these universities all exceeding 300 million Yuan (RMB) (US$49.14 million). Accordingly, there is a need for an in-depth study regarding the efficiency and RTS of the use of these funds for scientific research at these universities, a study that can serve as the basis for national macro-S&T management departments to make relevant decisions.

RTS is an important issue in the analysis of organisational performance in that it can help decision makers (DMs) decide if the size of the organisation should be expanded or reduced. RTS, a classic economic concept, is tied to the relationship between production factors and variations of outputs. If the scale of production changes due to the proportional increase (decrease) of all production factors, the RTS measures the proportion of the change rate of output(s) with that of all input(s) (Pindyck and Rubinfeld, 2000). There are three types of RTS in production processes. If proportional change of outputs is the same, less or more than that of inputs, we classify RTS as constant returns to scale (CRS), decreasing returns to scale (DRS) or increasing returns to scale (IRS), respectively. For universities, if IRS is found, they have the following options: (1) expand their research inputs (for example, staff and research funding) or (2) merge their research operations to increase scale efficiency. On the contrary, if DRS is found, they should (1) decrease their research inputs or (2) split their research operations into smaller units. However, if CRS is found, these universities are running on the optimal scale size.

The traditional definition of RTS in economics is based on the idea of measuring radial changes in outputs caused by those of all inputs. However, in some real applications, the increase in scale is often caused by the inputs changing in unequal proportions. Based on the above thinking, Yang and Liu (2017) introduce directional RTS from a global and local (directional scale elasticity) perspective under the Pareto preference and provide specific formulations of directional RTS. See also Yang (2012).

This chapter aims to estimate the directional RTS of the S&T activities of 38 Project 985 universities in China and addresses the following questions. Is the massive financial investment for S&T by the government being used efficiently? What are the efficiencies of S&T resource usage? Are these universities running on optimal scale? All of these issues relate to the rational allocation of the limited S&T resources and provide basic information for related S&T policies developed by the management teams of these universities and national S&T macro-management departments.

The chapter is organised as follows. Section 2 presents the methodology used in this chapter to analyse the directional RTS of the S&T activities of these universities using the methods proposed by Yang and Liu (2017) and Yang et al. (2019). In section 3, we discuss the results of the analyses, including the directional RTS and the congestion effect. The conclusions and discussions are provided in section 4.

2. METHODOLOGY

2.1 Input–Output Indicators

Performance management has become common in government-managed organisations in the past few years as a consequence of two principal factors: a) an increased demand for accountability by governing bodies, press and the general public and b) a growing commitment by organisations to focus on results to improve performance (Poister, 2003). The dictionary definition of 'performance' emphasises the situation of the organisation in the present or in the past, but not in the future. By this definition, 'performance' is observable and measurable − theoretically at least. However, economic 'performance' emphasises anticipated future results, which may be an estimation based on past or present behaviours of the organisation shown over a defined period of time. These two definitions reflect the two main stages of the theoretical development of performance measurement (Meyer, 2002). Meyer (2002 (page 30)) proposed seven purposes for these performance measures. He suggests that large and complex organisations require more from their measures than do smaller entities. The latter may only need measures that 'look ahead, look back, motivate and compensate people' whereas more complicated organisations require measures that 'roll up from the bottom to the top of the organisation, to cascade down from top to bottom, and to facilitate performance comparisons across business and functional units'.

The development of performance management during the past decade indicates a change from the output (result) model to the objective − process − result model (Zhang et al., 2011). Geisler (2000) defined metrics for evaluating scientific work as 'a system of measurement that includes: (1) the objective being measured; (2) the units to be measured; (3) the value of the units'. Keeney and Gregory (2005) studied how to select measures effectively to determine whether bodies that operate in such an environment are meeting their targets (that is, assessment indicators). Roper et al. (2004) discussed the indicators for the pre-assessment of public R&D funding based on the beneficial outcomes that increased knowledge provided to society. Moreover, Soft Systems Methodology (SSM) can be used to analyse systematically the operation of research institutions to build a relatively more complete and reasonable set of evaluation indicators based on the 3E theory (efficacy, efficiency, effectiveness) (Meng et al., 2007; Mingers et al., 2009). Zhang et al. (2011) proposed strategy maps for national research institutes (NRIs) based on discussions on the general rules of research activities so that managers can more clearly, accurately and logically describe the strategies of their organisations. In the real practice of analysing relative efficiencies and the RTS of universities so they can improve efficiencies of S&T resource utilisation and thereby play a more important role in the fields of economic development, social progress and national defence are important issues for the management of these universities. In the statistical practice of S&T activities of universities in China, dozens of quantitative indicators (for example, publications, awards, patents, staff, talents, funds, graduate training) are used to monitor the annual development status of these universities.

From the above studies, we note that there exist dozens of quantitative indicators for the inputs and outputs of the S&T activities of Project 985 universities. In this chapter, we use the indicators in 'S&T statistics compilation in 2017', which is published by the Ministry of Education in China, to analyse the directional RTS for the S&T activities of the Project 985 universities.

These indicators include: (1) input indicators: teaching and research staff (STAFF) and S&T funds (FUND); (2) output indicators: monographs, papers, technology transfer income (TT INCOME) and awards. As input indicators, STAFF refers to the employees registered in the universities in the statistical year who are engaged in teaching, research and development, applications of research and development results and scientific and technological services as well as those serving the above works, such as the foreign and domestic experts from outside the higher education system whose accumulated work time is more than one month. FUND denotes the total income of the universities, including research funding and block grants in the statistical year. As output indicators, monographs denote the academic works of a more comprehensive specialised topic of a subject. Papers denote the publications in important international and domestic journals in the statistical year. TT INCOME denotes the total income from the process of technology transfer in the university in the statistical year. Award refers to something given to a person or a group of people in the university to recognise their excellence in a certain field in the statistical year. Based on these indicators, we investigate the directional RTS of the S&T activities of 38 Project 985 universities in China in 2016.

2.2 Data

We provide the following descriptive statistics for the input–output data of the S&T activities of 38 Project 985 universities (denoted by DMU_1~DMU_{38}) in China in 2016. See Table 18.1. The list of names of those universities can also be found in Table 18.2. Detailed data are listed in Table 18.5 in section 3.2.

Table 18.1 Descriptive statistics for data of the S&T activities of 38 Project 985 universities in China in 2016

	Input indicators		Output indicators			
Variables	STAFF (Number)	FUND (RMB in million)	Monograph (Number)	Paper (Number)	TT income (RMB in million)	Award (Number)
Mean	4910.2895	1 929 039.2368	27.5000	7105.3421	29 575.2632	35.6053
Standard Deviation	3807.2960	1 204 489.5224	19.8425	4501.1753	63 782.5311	26.6503
Max	13 905.0000	5 168 392.0000	70.0000	20 701.0000	302 898.0000	112.0000
Min	249.0000	15 933.0000	0.0000	110.0000	0.0000	0.0000
Median	2779.5000	1 654 041.5000	27.0000	6373.5000	3756.0000	29.5000

Source: S&T statistics compilation in 2017.

Table 18.2 List of Project 985 universities

DMUs	Universities	DMUs	Universities
DMU_1	Beijing University	DMU_{20}	Southeast University
DMU_2	Renmin University of China	DMU_{21}	Zhejiang University
DMU_3	Tsinghua University	DMU_{22}	University of Science and Technology of China
DMU_4	Beihang University	DMU_{23}	Xiamen University
DMU_5	Beijing Institute of Technology	DMU_{24}	Shandong University
DMU_6	China Agricultural University	DMU_{25}	Ocean University of China
DMU_7	Beijing Normal University	DMU_{26}	Wuhan University
DMU_8	Minzu University of China	DMU_{27}	Huazhong University of Science and Technology
DMU_9	Nankai University	DMU_{28}	Hunan University
DMU_{10}	Tianjin University	DMU_{29}	Central South University
DMU_{11}	Dalian University of Technology	DMU_{30}	Sun Yat-sen University
DMU_{12}	Northeastern University	DMU_{31}	South China University of Technology
DMU_{13}	Jilin University	DMU_{32}	Chongqing University
DMU_{14}	Harbin Institute of Technology	DMU_{33}	Sichuan University
DMU_{15}	Fudan University	DMU_{34}	University of Electronic Science and Technology
DMU_{16}	Tongji University	DMU_{35}	Xi'an Jiaotong University
DMU_{17}	Shanghai Jiaotong University	DMU_{36}	Northwestern Polytechnical University
DMU_{18}	East China Normal University	DMU_{37}	Northwest A&F University
DMU_{19}	Nanjing University	DMU_{38}	Lanzhou University

The dataset comes from the official report, 'S&T statistics compilation in 2017', which is distributed throughout China. The report can be found in the main libraries throughout China, such as the National Science Library of the Chinese Academy of Sciences in Beijing and so on.

2.3 Existing Approaches for Directional RTS and Congestion

In the public sector, the production function often cannot be formulated as $F(Y, X) = 0$. Therefore, the data envelopment analysis (DEA) method is one of the most commonly used approaches for the analysis of RTS on the public sector (for example, research institutions) (Fox, 2002). The estimation of the RTS of DMUs using the DEA method was first investigated by Banker (1984) and Banker et al. (1984). Banker (1984) introduced the RTS to economics through the DEA framework and proposed the method to determine the RTS of DMUs in DEA models, which extended the application area of the DEA from relative efficiency evaluation to RTS measurement. He used the Charnes, Cooper and Rhodes (CCR)-DEA model with radial measure to estimate the DMUs' RTS. Soon after, Banker et al. (1984) proposed the Banker, Charnes and Cooper (BCC)-DEA model under the assumption of a variable RTS and investigated how to apply the BCC-DEA model to estimate the RTS of DMUs. Soleimani-damaneh (2012) discussed the PPS (production possibility set)-based definition of RTS, leading to a modified definition of RTS, which is suitable in the presence of multiple supporting hyperplanes passing through the unit under assessment. Thus far, an important group of the studies of DMUs' RTS are DEA based, and these research efforts can be roughly divided into two categories.

(1) RTS measurement using radial DEA models. The most well-known approach for determining the RTS of DMUs is to calculate the value of $\Sigma_{j=1}^{n} \lambda_j^*$ in CCR-DEA models, where λ_j^* denotes the jth weight of DMU_j. Banker et al. (1984) proposed the method to examine the intercept of the supporting hyperplane on the PPS under the variable RTS assumption. This intercept corresponds to a dual variable regarding the convex constraint in the BCC-DEA model. Färe and Grosskopf (1985) and Färe et al. (1983, 1985, 1994) used the Färe, Grosskopf and Lovell (FGL)-DEA model to measure the scale elasticity (SE). Førsund (1996) discussed the quantitative measurement of SE and RTS, which was extended further to mathematical characterisations of SE for both frontier and non-frontier units by Fukuyama (2000). Huang et al. (1997), Kerstens and Vanden Eeckaut (1998) and Read and Thanassoulis (2000) investigated the quantitative measurement of SE in DEA models. Research efforts related to this approach can be found in Chang and Guh (1991), Banker and Thrall (1992), Zhu and Shen (1995), Banker et al. (1996a, 1996b), Seiford and Zhu (1998, 1999), Tone (1996), Golany and Yu (1997), Sueyoshi (1999), Cooper et al. (2000), Tone and Sahoo (2003) and so on.

(2) RTS measurement using non-radial DEA models. Banker et al. (2004) discussed the RTS measurement using the additive model and multiplicative model. Sueyoshi and Sekitani (2005) explored the RTS measurement using dynamic DEA, a production scheme that includes a feedback process. Zarepisheh et al. (2010) discussed the RTS issue in multiplicative models, which is a single model in one stage that differs from the two-stage method proposed by Banker et al. (2004). Lozano and Gutierrez (2011) analysed the RTS of 41 Spanish airports using the DEA model with the Russell measure. Khodabakhshi et al. (2010) discussed the RTS issue in vague DEA models. Sueyoshi and Sekitani (2007) theoretically explored the measurement of RTS using a non-radial model with a range-adjusted measure. A new linear programming RAM/RTS approach was proposed to address the simultaneous occurrence of multiple reference sets, multiple supporting hyperplanes and multiple projections. Soleimani-damaneh et al. (2006) explored the RTS measurement in the FDH model. Fukuyama (2003) studied RTS in the DEA framework in the context of the directional technology distance function. In fact, in the RTS measurement, the projections on the efficient frontier of DMUs within the PPS are different between radial DEA models and non-radial DEA models, but this difference does not affect the RTS of DMUs on the efficient frontier.

The existing RTS measurements in DEA models are all based on the definition of RTS in the DEA framework of Banker (1984). It is noted that the directional RTS in the DEA framework in this chapter is very different from those of classic RTSs using either radial or non-radial DEA models. The traditional definition of RTS adopted in Banker's framework is based on the notion that the measure of radial or non-radial changes in outputs is the result of the **radial changes** of all inputs (that is, all components of inputs change in the same proportion). Yang and Liu (2017) argued that, due to the complexity of research activities in research institutions, it can often be observed that production factors are not necessarily tied together proportionally, and inputs often change non-proportionally. Based on this observation, Yang and Liu (2017) introduced directional RTS from a global and local (directional scale elasticity) perspective under the Pareto preference and provides specific formulations of directional RTS assuming that the inputs change in different proportions. In addition, Yang and Liu (2017) demonstrated that traditional RTS is a special case of directional RTS with a radial direction, such that directional RTS can provide a basis for decisions regarding the further development of such production processes.

In the process of estimating RTS based on the production function or the PPS, a congestion effect is often involved. Essentially, the congestion effect describes the issue of excessive inputs (Wei and Yan, 2004). A congestion effect means the reduction of one (or some) input(s) will result in the increase of possibly one (or some) output(s) under the premise that other inputs or outputs do not deteriorate (Cooper et al., 2004). Färe and Grosskopf (1983, 1985) investigated the congestion effect using quantitative methods and proposed corresponding DEA models to address this issue. Soon after, Cooper et al. (1996) proposed another model to study the congestion effect. Cooper et al. (2001) compared the similarities and differences of the above two models. Wei and Yan (2004) and Tone and Sahoo (2004) built a new DEA model based on the new PPS under the assumption of weak disposal to detect the congestion effect of DMUs, which is denoted by the Wei-Yan-Tone-Sahoo (WY-TS) model. The above methods for congestion are based on the premise of radial changes in all inputs and outputs.

In this chapter, we will analyse the directional RTS of these universities to determine whether and how the inputs of these DMUs are to be expanded or further reduced. In particular, we determine whether there exists a directional congestion effect. If a directional congestion effect exists, we should not increase inputs of DMUs in the same direction as the congestion effect.

2.3.1 Directional RTS

The definition of directional scale elasticity in economics in the case of multiple inputs and multiple outputs proposed by Yang and Liu (2017) is $e\left(Y_0, X_0\right) = \left(-\Sigma_{i=1}^{m} \frac{\partial F}{\partial x_i} x_i \omega_i \Big/ \Sigma_{r=1}^{S} \frac{\partial F}{\partial y_r} y_r \delta_r\right)\Big|$ $\left(Y_0, X_0\right)$, where $F(Y, X) = 0$ denotes the production function and $e\left(Y_0, X_0\right)$ stands for the scale elasticity at the point of $\left(Y_0, X_0\right)$. Besides, we suppose the production function is continuously differentiable in the directions of $\left(\omega_1, \omega_2, ..., \omega_m\right)^T$ and $\left(\delta_1, \delta_2, ..., \delta_s\right)^T$, and $\Sigma_{r=1}^{S} \delta_r = s$ and $\Sigma_{i=1}^{m} \omega_i = m$ hold. The traditional scale elasticity only considers the elasticity for the input change along the diagonal directional, while ours considers all possible directions.

Remark 18.1. *In many applications, the above formula may not hold. In such a case, the differential in the above formula must be replaced by the left-hand directional derivative* ($t \to 0^-$) *and the right-hand directional derivative* ($t \to 0^+$) *such that we will have the left-hand and right-hand scale elasticities (see Banker, 1984; Podinovski and Førsund, 2010; Atici and Podinovski, 2012).*

Suppose there are n DMUs to be evaluated that are indexed by $j = 1, \ldots, n$ and each DMU is assumed to produce s different outputs from m different inputs, donated, respectively, by $x_j = (x_{1j}, \ldots, x_{mj})$ and $y_j = (y_{1j}, \ldots, x_{sj})$. Further, assume that all components of vectors x_j and y_j for all DMUs are non-negative, and at the same time, each DMU has at least one strictly positive input and one strictly positive output. We assume $y_{r0}(r = 1, 2, \ldots, s)$ and $x_{i0}(i = 1, 2, \ldots, m)$

as the output and input values of DMU_0, which is the DMU to be evaluated and denoted by $DMU(Y_0, X_0)$. Based on the above definition of directional scale elasticity, Yang and Liu (2017) proposed the definition of directional RTS in a DEA framework based on the production possibility set (PPS) as follows.

Definition 18.1. Assuming $DMU(X_0, Y_0) \in PPS$ and $X_0 \in R_m^+, Y_0 \in R_s^+$, the directional RTS can be defined as $\beta(t) = \max\{\beta | (\Omega_t X_0, \Phi_\beta Y_0) \in PPS \text{ and } t \neq 0\}$.

Therein, $\Omega_t = diag\{1+\omega_1 t, ..., 1+\omega_m t\}$ and $\Phi_\beta = diag\{1+\delta_1\beta, ..., 1+\delta_s\beta\}$, $diag\{\bullet\}$ denotes the diagonal matrix, $(\omega_1, ..., \omega_m)^T$ $(\omega_i \geq 0, i=1, ..., m)$ and $(\delta_1, ..., \delta_s)^T$ $(\delta_r \geq 0, r=1, ..., s)$ represent input and output directions, respectively, and satisfy $\Sigma_{i=1}^m \omega_i = m$ and $\Sigma_{r=1}^S \delta_r = s$ where t, β are input and output scaling factors, respectively.

If we further set the following definition of the limit function for the directional RTS,

$$\rho^- = \lim_{t \to 0^-} \frac{\beta(t)}{t} \tag{18.1}$$

$$\rho^+ = \lim_{t \to 0^+} \frac{\beta(t)}{t} \tag{18.2}$$

then, we have

(a) if $\rho^- > 1$ (or $\rho^+ > 1$) holds, then increasing directional RTS prevails on the left-hand (or right-hand) side of this point (X_0, Y_0) in the direction of $(\omega_1, \omega_2, . . ., \omega_m)^T$ and $(\delta_1, \delta_2, . . ., \delta_s)^T$;
(b) if $\rho^- = 1$ (or $\rho^+ = 1$) holds, then constant directional RTS prevails on the left-hand (or right-hand) side of this point (X_0, Y_0) in the direction of $(\omega_1, \omega_2, . . ., \omega_m)^T$ and $(\delta_1, \delta_2, . . ., \delta_s)^T$;
(c) if $\rho^- < 1$ (or $\rho^+ < 1$) holds, then decreasing directional RTS prevails on the left-hand (or right-hand) side of this point (X_0, Y_0) in the direction of $(\omega_1, \omega_2, . . ., \omega_m)^T$ and $(\delta_1, \delta_2, . . ., \delta_s)^T$.

Definition 18.2. The weakly and strongly efficient frontiers of PPS are defined as follows.
(1) Weakly efficient frontier:

$$EF_{weak} = \{(X, Y) \in PPS \,|\, \text{there is no } (\overline{X}, \overline{Y}) \in PPS \text{ such that } (-X, Y) < (-\overline{X}, \overline{Y})\} \tag{18.3a}$$

(2) Strongly efficient frontier:

$$EF_{strong} = \left\{ \begin{array}{l} (X, Y) \in PPS \,|\, \text{there is no } (\overline{X}, \overline{Y}) \in PPS \\ \text{such that } (-X, Y) \leq (-\overline{X}, \overline{Y}) \text{ and } (\overline{X}, \overline{Y}) \neq (X, Y) \end{array} \right\} \tag{18.3b}$$

The following method is used to estimate the directional RTS based on **Definition 18.1** (Yang and Liu 2017). For the strongly efficient DMU(X_0, Y_0) on the strongly efficient frontier in the BCC-DEA model, its directional scale elasticity is determined using Model (18.4):

$$\bar{\rho}\left(\underline{\rho}\right)=\max\left(\min\right)\frac{V^TWX_0}{U^T\Delta Y_0}\ s.t.\begin{cases} U^TY_j-V^TX_j+\mu_0\le 0,\ j=1,...,n \\ U^TY_0-V^TX_0+\mu_0=0 \\ V^TX_0=1 \\ U\ge\mathbf{0},V\ge\mathbf{0},\mu_0\ \text{free}\end{cases} \tag{18.4}$$

where $U=(u_1, u_2, \ldots, u_s)^{\mathrm{T}}$ and $V=(v_1, v_2, \ldots, v_m)^T$ are vectors of multipliers and $\Delta = diag\ \{\delta_1, \delta_2, \ldots, \delta_s\}$ and $W = diag\ \{\omega_1, \omega_2, \ldots, \omega_m\}$ are matrixes of input and output directions, respectively.

Based on the optimal solutions of Model (18.4), we employ the following procedure for determining the directional RTS of DMU (X_0, Y_0), which is a strongly efficient DMU on the strongly efficient frontier of PPS in the BCC-DEA model (denoted as $P_{BCC}(X, Y)^1$), in the direction of $(\omega_1, \ldots, \omega_m)^T$ and $(\delta_1, \ldots, \delta_s)^T$.

(1) The directional RTS to the right of DMU(X_0, Y_0):

  (a) $\underline{\rho}\left(X_0,Y_0\right)>1$, increasing directional RTS prevails;
  (b) $\underline{\rho}\left(X_0,Y_0\right)=1$, constant directional RTS prevails;
  (c) $\underline{\rho}\left(X_0,Y_0\right)<1$, decreasing directional RTS prevails.

(2) The directional RTS to the left of DMU(X_0, Y_0):

  (a) $\bar{\rho}\left(X_0,Y_0\right)>1$, increasing directional RTS prevails;
  (b) $\bar{\rho}\left(X_0,Y_0\right)=1$, constant directional RTS prevails;
  (c) $\bar{\rho}\left(X_0,Y_0\right)<1$, decreasing directional RTS prevails;
  (d) if the optimal objective value $\bar{\rho}\left(X_0,Y_0\right)$ of Model (18.4) is unbounded ($+\infty$), there are no data to determine the directional RTS to the left of DMU(X_0, Y_0).

For inefficient or weakly efficient DMUs on a weakly efficient frontier, we project them onto the strongly efficient frontier using DEA models and estimate the directional RTS to the right and left according to the directional RTS of these projections.

Obviously, as Model (18.4) is fractional programming, it is difficult to solve. Therefore, we transform it into equivalent mathematical programming (Model 5) through a Charnes-Cooper transformation (Charnes and Cooper, 1962).

$$\bar{\rho}\left(\underline{\rho}\right)=\max\left(\min\right)\Gamma^TX_0\ \ s.t.\begin{cases} \Lambda^T\Delta^{-1}Y_j-\Gamma^TW^{-1}X_j+\mu_0'\le 0,\ j=1,...,n \\ \Lambda^T\Delta^{-1}Y_0-\Gamma^TW^{-1}X_0+\mu_0'=0 \\ \Gamma^TW^{-1}X_0=t \\ \Lambda^TY_0=1 \\ \Gamma\ge\mathbf{0},\Lambda\ge\mathbf{0},t\ge 0,\mu_0'\ \text{free}\end{cases} \tag{18.5}$$

We obtain the directional scale elasticity, the directional RTS of DMU(X_0, Y_0) and its optimal input direction by solving Model (18.5).

2.3.2 Directional congestion

Yang and Liu (2017) argued that, due to the complexity of research activities in research institutions, it can often be observed that production factors are not necessarily tied together proportionally, and inputs change non-proportionally. Accordingly, Yang et al. (2019) introduced the concept of directional congestion under the Pareto preference and offer specific formulations and models. The methods they proposed are as follows.

For the strongly efficient DMU(X_0, Y_0) on the strongly efficient frontier of the production possibility set (denoted by P_{convex} $(X, Y)^2$) determined in Model (18.6):

$$\max \theta = \theta_0 \quad s.t.\begin{cases} \sum_j \lambda_j x_{ij} = x_{i0}, i = 1, ..., m \\ \sum_j \lambda_j y_{rj} - s_r^+ = \theta_0 y_{r0}, r = 1, ..., s \\ \sum_j \lambda_j = 1, \quad j = 1, ..., n \\ \lambda_j, s_r^+ \geq 0 \end{cases} \tag{18.6}$$

The directional scale elasticity of DMU(X_0, Y_0) can be determined through the following Model (18.7):

$$\bar{\rho}(\underline{\rho}) = \max(\min) \frac{V^T W X_0}{U^T \Delta Y_0} \quad s.t.\begin{cases} U^T Y_j - V^T X_j + \mu_0 \leq 0, j = 1, ..., n \\ U^T Y_0 - V^T X_0 + \mu_0 = 0 \\ V^T X_0 = 1 \\ U \geq \mathbf{0}, V, \mu_0 \text{ free} \end{cases} \tag{18.7}$$

We transform Model (18.7) into an equivalent mathematical programming (Model 18.8) through Charnes-Cooper transformation (Charnes and Cooper, 1962) as follows.

$$\bar{\rho}(\underline{\rho}) = \max(\min) \Gamma^T X_0 \quad s.t.\begin{cases} \Lambda^T \Delta^{-1} Y_j - \Gamma^T W^{-1} X_j + \mu_0' \leq 0, j = 1, ..., n \\ \Lambda^T \Delta^{-1} Y_0 - \Gamma^T W^{-1} X_0 + \mu_0' = 0 \\ \Gamma^T W^{-1} X_0 = t \\ \Lambda^T Y_0 = 1 \\ \Lambda \geq \mathbf{0}, t \geq 0, \Gamma, \mu_0' \text{ free} \end{cases} \tag{18.8}$$

Based on the results of Model (18.8), we apply the following procedure for determining the congestion effect of strongly efficient DMU(X_0, Y_0) on strongly efficient frontier of $P_{convex}(X, Y)$ in the direction of $(\omega_1, \ldots, \omega_m)^T$ and $(\delta_1, \ldots, \delta_s)^T$.

(1) If the optimal objective value $\underline{\rho}(X_0, Y_0)$ in Model (18.8) is bounded and satisfies $\underline{\rho}(X_0, Y_0) < 0$, the directional congestion effect occurs to the right of DMU(X_0, Y_0). Otherwise, if the optimal objective value $\underline{\rho}(X_0, Y_0)$ in Model (18.8) is unbounded (-∞), there are no data to determine the directional congestion effect to the 'right' of DMU(X_0, Y_0).
(2) If the optimal objective value $\overline{\rho}(X_0, Y_0)$ in Model (18.8) is bounded and satisfies $\overline{\rho}(X_0, Y_0) < 0$, the directional congestion effect occurs to the left of DMU(X_0, Y_0). Otherwise, if the optimal objective value $\overline{\rho}(X_0, Y_0)$ in Model (18.8) is unbounded (+∞), there are no data to determine the directional congestion effect to the 'left' of DMU(X_0, Y_0).

We project inefficient or weakly efficient DMUs onto the strongly efficient frontier using DEA models such that we can detect the directional congestion effect to the right and left according to the projections.

3. ANALYSIS RESULTS OF DIRECTIONAL RTS AND DIRECTION CONGESTION EFFECT

3.1 Directional RTS

First, we determine the strongly efficient frontier using the input-based BCC-DEA model (Model 18.9) with radial measurement.

$$\min \theta = \theta_0 - \varepsilon\left(\sum_r s_r^+ + \sum_i s_i^-\right) s.t. \begin{cases} \sum_j \lambda_j x_{ij} + s_i^- = \theta_0 x_{i0}, i = 1, ..., m \\ \sum_j \lambda_j y_{rj} - s_r^+ = y_{r0}, r = 1, ..., s \\ \sum_j \lambda_j = 1, \lambda_j, s_i^-, s_r^+ \geq 0 \end{cases} \tag{18.9}$$

Using Model (18.9), we obtain the efficiencies and slacks of the S&T activities of 38 Project 985 universities in China in 2016. We obtain the projections of these universities on the strongly efficient frontier (See Table 18.3) using formula (18.10).

$$\tilde{x}_{i0} \leftarrow \theta_0^* x_{i0} - s_i^{-*}, i = 1, ..., m\ ;\ \tilde{y}_{r0} \leftarrow y_{r0} + s_r^{+*}, r = 1, ..., s \tag{18.10}$$

Second, we determine the directional RTS of the S&T activities of 38 projected Project 985 universities in China in 2016 using Model (18.5) in section 2. We use DMU_4 as an example because it offers interesting results. Without loss of generality, we set the output directions as $\delta_1 = \delta_2 = \delta_3 = \delta_4 = 1$. See Table 18.4 and Figures 18.1−18.2 for details.

Table 18.3 Projections of the S&T activities of 38 Project 985 universities on the strong efficient frontier produced by Model (18.9)

Universities	Input indicators		Output indicators			
	STAFF (Number)	FUND (RMB in million)	Monograph (Number)	Paper (Number)	TT income (RMB in million)	Award (Number)
DMU_1	10 343	2 666 622	56	15 489	14 445	90
DMU_2	227	85 145	10	725	10 490	6
DMU_3	5583	5 168 392	63	9016	302 898	86
DMU_4	2315	2 701 240	8	7384	2576	24
DMU_5	2439	2 625 658	41	6305	253 765	28
DMU_6	1597	1 263 808	35	3520	22 396	27
DMU_7	1251	628 449	29	2197	17 235	21
DMU_8	249	15 933	4	622	0	0
DMU_9	1913	608 330	8	3521	29 156	12
DMU_{10}	3049	2 514 015	12	8527	36 653	14
DMU_{11}	2877	1 389 060	33	5954	500	53
DMU_{12}	2325	1 607 640	70	4094	48 045	49
DMU_{13}	7578	1 306 589	42	10 380	7356	86
DMU_{14}	3451	2 439 908	30	9367	566	40
DMU_{15}	6872	2 424 153	41	12 529	65 034	42
DMU_{16}	3200	1 558 495	33	6843	8806	41
DMU_{17}	13 333	3 655 869	56	20701	12 574	88
DMU_{18}	874	416 436	11	2088	2516	9
DMU_{19}	1799	1 086 248	32	3744	13 961	32
DMU_{20}	3378	1 375 889	38	6442	20 995	36
DMU_{21}	7085	2 787 716	50	11 103	101 208	83
DMU_{22}	1583	1 632 785	7	4853	6400	14
DMU_{23}	1367	643 471	16	2964	267	22
DMU_{24}	5683	1 084 113	30	8462	15 300	45
DMU_{25}	1833	637 587	13	3904	1950	17
DMU_{26}	7926	1 447 858	52	10 575	7400	104
DMU_{27}	9368	3 151 857	64	11 745	14 972	51
DMU_{28}	1183	512 312	11	2700	2650	13
DMU_{29}	8751	1 439 111	50	11 539	1980	112
DMU_{30}	6615	2 730 226	44	12 463	11 129	57
DMU_{31}	2645	2 409 116	27	7101	89 671	33
DMU_{32}	2569	820 863	16	5272	712	24
DMU_{33}	9899	2 139 423	46	15 777	47 223	42

(*Continued*)

Table 18.3 (Continued)

Universities	Input indicators: STAFF (Number)	FUND (RMB in million)	Output indicators: Monograph (Number)	Paper (Number)	TT income (RMB in million)	Award (Number)
DMU_{34}	1948	891 304	27	3980	12 262	25
DMU_{35}	4658	1 097 547	28	7220	36 900	38
DMU_{36}	1966	1 801 188	32	4901	20 500	30
DMU_{37}	2476	566 200	17	4108	10 153	16
DMU_{38}	1422	306 200	10	2503	311	14

Table 18.4 Directional RTS of the S&T activities of DMU_4 in China in 2016

DMU	ω_1	ω_2	$\underline{\rho}$ (Lower bound)	$\overline{\rho}$ (Upper bound)	Directional RTS (right)	Directional RTS (left)
	0.1	1.9	0.05	0.76	Decreasing	Decreasing
	0.2	1.8	0.11	0.78	Decreasing	Decreasing
	0.3	1.7	0.16	0.81	Decreasing	Decreasing
	0.4	1.6	0.22	0.83	Decreasing	Decreasing
	0.5	1.5	0.27	0.86	Decreasing	Decreasing
	0.6	1.4	0.33	0.88	Decreasing	Decreasing
	0.7	1.3	0.38	0.91	Decreasing	Decreasing
	0.8	1.2	0.44	0.93	Decreasing	Decreasing
	0.9	1.1	0.49	0.96	Decreasing	Decreasing
DMU_4	1	1	0.54	1.05	Decreasing	Increasing
	1.1	0.9	0.60	1.16	Decreasing	Increasing
	1.2	0.8	0.65	1.27	Decreasing	Increasing
	1.3	0.7	0.71	1.37	Decreasing	Increasing
	1.4	0.6	0.76	1.48	Decreasing	Increasing
	1.5	0.5	0.82	1.58	Decreasing	Increasing
	1.6	0.4	0.87	1.69	Decreasing	Increasing
	1.7	0.3	0.93	1.79	Decreasing	Increasing
	1.8	0.2	0.98	1.90	Decreasing	Increasing
	1.9	0.1	1.03	2.00	Increasing	Increasing

In columns 2 and 3 of Table 18.4, it is evident that the right-hand and left-hand directional RTSs of DMU_4 are in different input directions. The lower and upper bounds of Model (18.5) are presented in columns 4 and 5. According to the procedure for determining the directional RTS in section 2.3.1, we know the types of directional RTS in different input directions. Based

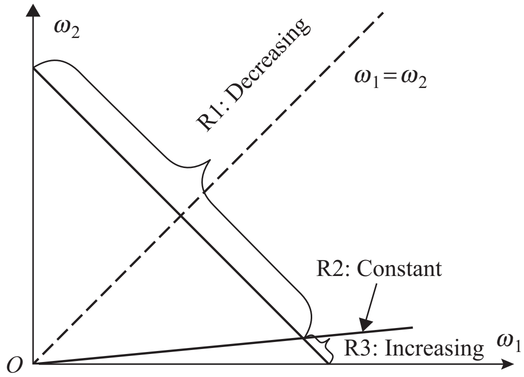


Figure 18.1 Directional RTS to the right of DMU_4

on the types of directional RTS in different input directions, we can illustrate the types of directional RTS of DMU_4 using Figures 18.1−18.2.

The above analysis yields the following findings. On the basis of existing inputs, if STAFF and FUND increase with the proportion located in R1, decreasing directional RTS prevails, otherwise constant or increasing directional RTS prevails. If STAFF and FUND decrease in radial proportion, decreasing directional RTS prevails. If the proportion of STAFF and FUND is located in the R1 area in Figure 18.2, increasing directional RTS prevails. If the proportion of input decrease is located in the R3 area, decreasing directional RTS prevails. If the proportion of input decrease is located in the R2 area, constant directional RTS prevails. See Figure 18.2.

Similarly, we have the directional RTS to the 'right' and 'left' of other DMUs. The directional RTS of the S&T activities of these 38 Project 985 universities can serve as one of the bases for decisions regarding organisational adjustments for these universities.

3.2 Directional Congestion Effect

As previously mentioned herein, a congestion effect describes the issue of excessive inputs (Wei and Yan, 2004). In this case, an increase (decrease) in one or more inputs causes a

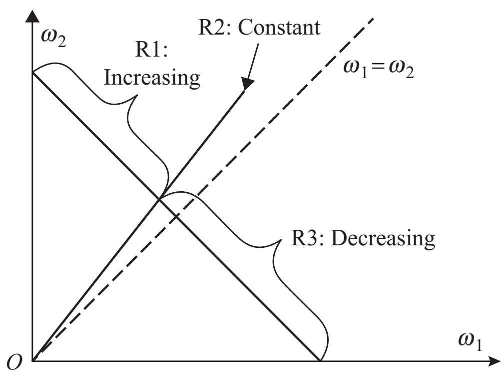


Figure 18.2. Directional RTS to the left of DMU_4

decrease (increase) in one or more outputs (Cooper et al., 2001). For a university, if there exists a congestion effect, its inputs should be decreased so that its outputs can increase. This information is also important for DMs when determining the S&T resource allocation. Therefore, in this section, we determine whether there exists a congestion effect. Furthermore, if a congestion effect exists, we identify the directions of the congestion effect so we can assess whether an increase (decrease) in multiple inputs of these universities in a specific direction causes the decrease (increase) in one or more outputs.

We first detect the congestion effect of the S&T activities of 38 Project 985 universities using the WY-TS model (Wei and Yan, 2004; Tone and Sahoo, 2004; Kao, 2010) based on the input–output data of these universities. We use the following model to measure pure technical efficiency (Tone and Sahoo, 2004; Kao, 2010):

$$\max \eta = \eta_0 \quad s.t. \begin{cases} \sum_{j=1}^{n} \lambda_j x_{ij} = x_{i0},\, i = 1, ..., m \\ \sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = \eta_0 y_{r0},\, r = 1, ..., s \\ \sum_{j=1}^{n} \lambda_j = 1,\, \lambda_j, s_r^+ \geq 0 \end{cases} \tag{18.11}$$

According to Model (18.11), we obtain the efficiencies and slacks of the S&T activities of 38 Project 985 universities in China in 2016. We use the following Model (18.12) to obtain the projections of the S&T activities of 38 Project 985 universities on the strongly efficient frontier of $P_{convex}(X, Y)$ produced by Model (18.11) (Wei and Yan, 2004; Tone and Sahoo, 2004). See Table 18.5 for details.

$$\begin{cases} \tilde{x}_{i0} \leftarrow x_{i0},\, i = 1, ..., m \\ \tilde{y}_{r0} \leftarrow \eta_0^* y_{r0} + s_r^{+*},\, r = 1, ..., s \end{cases} \tag{18.12}$$

We further put the corresponding projections of these DMUs into the original dataset, and we then obtain the efficiencies and slacks of their projections using the output-based BCC-DEA model in Model (18.13).

$$\max \theta = \theta_0 + \varepsilon \sum_r s_r^+ \quad s.t. \begin{cases} \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{i0},\, i = 1, ..., m \\ \sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = \theta_0 y_{r0},\, r = 1, ..., s \\ \sum_{j=1}^{n} \lambda_j = 1,\, \lambda_j, s_i^-, s_r^+ \geq 0 \end{cases} \tag{18.13}$$

where $\varepsilon > 0$ is a non-Archimedean construct. Note that Model (18.13) differs from Model (18.11) only in the first constraint set where $\sum_j \lambda_j x_{ij} + s_i^- = x_{i0},\, i = 1, ..., m$ is replaced by $\sum_j \lambda_j x_{ij} = x_{i0},\, i = 1, ..., m$. The equality constraints in Model (18.13) are stronger than the inequality constraints of Model (18.11).

As proposed by Tone and Sahoo (2004) and Kao (2010), the congestion effect can be reflected in the results of θ^*, since it not only reveals the improvement from θ_0 but also from

Table 18.5 Projections of the S&T activities of 38 Project 985 universities on the strong efficient frontier produced by Model (18.12)

Universities	Input indicators		Output indicators			
	STAFF (Number)	FUND (RMB in million)	Monograph (Number)	Paper (Number)	TT income (RMB in million)	Award (Number)
DMU_1	13 905	2 846 243	56	15 489	14 445	29
DMU_2	270	85 196	1	110	0	2
DMU_3	5583	5 168 392	63	9016	302 898	86
DMU_4	2315	2 701 240	8	7384	2576	24
DMU_5	2439	2 625 658	41	6305	253 765	28
DMU_6	1957	1 684 420	38	4431	25 086	33
DMU_7	1424	680 526	31	2434	18 160	24
DMU_8	249	15 933	4	622	0	0
DMU_9	1913	608 330	8	3521	29 156	12
DMU_{10}	3049	2 514 015	12	8527	36 653	14
DMU_{11}	2877	1 389 060	33	5954	500	53
DMU_{12}	2325	1 607 640	70	4094	48 045	49
DMU_{13}	9417	1 623 663	51	12 073	3890	101
DMU_{14}	3451	2 439 908	30	9367	566	40
DMU_{15}	10 147	3 579 478	53	16 416	85 211	75
DMU_{16}	6037	2 940 106	52	10 736	66 353	64
DMU_{17}	13 333	3 655 869	56	20 701	12 574	88
DMU_{18}	1229	585 876	15	2808	4191	13
DMU_{19}	2182	1 317 428	36	4586	13 794	39
DMU_{20}	5030	2 048 545	54	9105	34 427	47
DMU_{21}	11 232	4 419 636	45	11 103	101 208	83
DMU_{22}	2571	2 651 307	11	7787	10 269	24
DMU_{23}	2275	1 070 953	24	4876	418	36
DMU_{24}	7973	1 459 066	47	10 732	18 755	102
DMU_{25}	2305	801 900	15	4857	2426	22
DMU_{26}	10 966	1 887 205	52	10 575	5785	95
DMU_{27}	9368	3 151 857	64	11 745	14 972	51
DMU_{28}	2037	882 450	18	4484	4401	22
DMU_{29}	8751	1 439 111	50	11 539	1980	112
DMU_{30}	8410	3 902 831	53	14 088	110 234	76
DMU_{31}	2682	2 525 034	27	7161	90 434	33
DMU_{32}	2569	820 863	16	5272	712	24
DMU_{33}	9899	2 139 423	46	15 777	47 223	42

(*Continued*)

Table 18.5 (Continued)

Universities	Input indicators: STAFF (Number)	FUND (RMB in million)	Output indicators: Monograph (Number)	Paper (Number)	TT income (RMB in million)	Award (Number)
DMU_{34}	2332	1 066 965	32	4705	14 923	30
DMU_{35}	6909	1 560 242	45	9755	49 855	86
DMU_{36}	2386	2 334 111	33	6160	26 123	35
DMU_{37}	3035	694 164	20	4933	13 614	19
DMU_{38}	1759	378 847	12	3015	1148	17

the slacks $s_r^+ (r = 1, ..., s)$. If $\theta^* = 1$, a congestion effect does not exist. If $\theta^* > 1$, a congestion effect exists.

After calculation, the results are presented in Table 18.6. The results in this table only detect congestion when the inputs increase proportionally. Thus, we find that the congestion effect occurs on DMU_1, DMU_6, DMU_{10}, DMU_{13}, DMU_{15}, DMU_{16}, DMU_{18}, DMU_{19}, DMU_{21}, DMU_{22}, DMU_{26}, DMU_{27}, DMU_{28}, DMU_{30}, DMU_{31}, DMU_{35} and DMU_{36}. We can further project these DMUs onto the strongly efficient frontier of $P_{BCC}(X, Y)$ produced from Model (18.13) by θ^* and their corresponding slacks.

Table 18.6 Congestion effect of the S&T activities of 38 Project 985 universities using WY-TS model

Universities	Input indicators: STAFF (Number)	FUND (RMB in million)	Output indicators: Monograph (Number)	Paper (Number)	TT income (RMB in million)	Award (Number)	Congestion effect (WY-TS model)
DMU_1	13 905	2 846 243	56	15 489	14 445	29	2220.35
DMU_2	270	85 196	1	110	0	2	1.00
DMU_3	5583	5 168 392	63	9016	302 898	86	1.00
DMU_4	2315	2 701 240	8	7384	2576	24	1.00
DMU_5	2439	2 625 658	41	6305	253 765	28	1.00
DMU_6	1957	1 684 420	17	3520	9340	27	819.08
DMU_7	1424	680 526	29	2197	650	6	1.00
DMU_8	249	15 933	4	622	0	0	1.00
DMU_9	1913	608 330	8	3521	29 156	12	1.00
DMU_{10}	3049	2 514 015	12	8527	36 653	14	1.00
DMU_{11}	2877	1 389 060	33	5954	500	53	1.00
DMU_{12}	2325	1 607 640	70	4094	48 045	49	1.00
DMU_{13}	9417	1 623 663	18	10 380	2280	59	8.29

Table 18.6 (Continued)

Universities	Input indicators		Output indicators				Congestion effect (WY-TS model)
	STAFF (Number)	FUND (RMB in million)	Monograph (Number)	Paper (Number)	TT income (RMB in million)	Award (Number)	
DMU_{14}	3451	2 439 908	30	9367	566	40	1.00
DMU_{15}	10 147	3 579 478	40	12 529	65 034	34	1.00
DMU_{16}	6037	2 940 106	33	6843	1098	41	1.00
DMU_{17}	13 333	3 655 869	56	20 701	12 574	88	1.00
DMU_{18}	1229	585 876	11	2088	512	6	1.00
DMU_{19}	2182	1 317 428	29	3744	3349	32	1.02
DMU_{20}	5030	2 048 545	38	6442	4258	26	1.00
DMU_{21}	11 232	4 419 636	45	11 103	101 208	83	5.33
DMU_{22}	2571	2 651 307	7	4853	6400	9	1.22
DMU_{23}	2275	1 070 953	14	2964	200	22	1.00
DMU_{24}	7973	1 459 066	4	8462	15 300	37	1.00
DMU_{25}	2305	801 900	0	3904	1950	17	1.00
DMU_{26}	10 966	1 887 205	52	10 575	5785	95	23 093.95
DMU_{27}	9368	3 151 857	64	11 745	14 972	51	1.00
DMU_{28}	2037	882 450	11	2700	2650	13	1.00
DMU_{29}	8751	1 439 111	50	11 539	1980	112	1.00
DMU_{30}	8410	3 902 831	44	12 463	2019	43	1.00
DMU_{31}	2682	2 525 034	12	7101	89 671	33	2.52
DMU_{32}	2569	820 863	16	5272	712	24	1.00
DMU_{33}	9899	21 39 423	46	15 777	47 223	42	1.00
DMU_{34}	2332	1 066 965	27	3980	4163	18	1.00
DMU_{35}	6909	1 560 242	27	7220	36 900	38	1.00
DMU_{36}	2386	2 334 111	12	4901	1910	30	1.00
DMU_{37}	3035	694 164	17	4108	3018	16	1.00
DMU_{38}	1759	378 847	0	2503	100	14	1.00

Source: S&T statistics compilation in 2017.

Second, we analyse the directional congestion effect of 38 Project 985 universities using the method (Model 18.8) mentioned in section 2.3.2 (directional congestion effect). We use DMU_{33} as an example because it offers interesting results. Without loss of generality, we set the outputs direction as $\delta_1 = \delta_2 = \delta_3 = \delta_4 = 1$, and we determine the directional congestion effect of this DMU in different input directions. See Table 18.7 for details.

Based on the above analysis, we find that congestion effect does not occur on DMU_{33} when using the WY-TS model. However, we also note that directional congestion effect occurs to

Table 18.7 Directional congestion effect of DMU_{33} in different inputs directions

DMU	ω1	ω2	$\underline{\rho}$	$\overline{\rho}$	Directional congestion effect (right)	Directional congestion effect (left)
	0.1	1.9	unbounded	−6.2547	No	Yes
	0.2	1.8	0.2777	unbounded	No	No
	0.3	1.7	unbounded	−4.9571	No	Yes
	0.4	1.6	unbounded	−4.3083	No	Yes
	0.5	1.5	0.3395	−3.6594	No	Yes
	0.6	1.4	0.2716	−3.0106	No	Yes
	0.7	1.3	0.2037	unbounded	No	No
	0.8	1.2	0.1358	unbounded	No	No
	0.9	1.1	unbounded	unbounded	No	No
DMU_{33}	1.0	1.0	0.0000	−0.9816	No	Yes
	1.1	0.9	−0.6729	−1.0047	Yes	Yes
	1.2	0.8	−1.3457	unbounded	Yes	No
	1.3	0.7	unbounded	unbounded	No	No
	1.4	0.6	unbounded	−1.1324	No	Yes
	1.5	0.5	unbounded	unbounded	No	No
	1.6	0.4	−4.0371	unbounded	Yes	No
	1.7	0.3	unbounded	unbounded	No	No
	1.8	0.2	unbounded	unbounded	No	No
	1.9	0.1	6.0557	unbounded	Yes	No

the right of DMU_{33} in certain directions (for example, $\omega_1 = 1.9$, $\omega_2 = 0.1$; $\delta_1 = \delta_2 = \delta_3 = \delta_4 = 1$) and occurs to the left of this DMU in some other directions (for example, $\omega_1 = 0.1$, $\omega_2 = 1.9$; $\delta_1 = \delta_2 = \delta_3 = \delta_4 = 1$). The reason for this phenomenon should be thoroughly investigated. Similarly, we analyse the directional congestion effect for other DMUs.

3.3 Policy Implications

From the above analysis results, we conclude that (1) the regions of increasing (constant, decreasing) directional RTS can be detected; (2) the congestion effect occurs on DMU_1, DMU_6, DMU_{10}, DMU_{13}, DMU_{15}, DMU_{16}, DMU_{18}, DMU_{19}, DMU_{21}, DMU_{22}, DMU_{26}, DMU_{27}, DMU_{28}, DMU_{30}, DMU_{31}, DMU_{35} and DMU_{36} when using the traditional WY-TS model. However, in certain directions, the directional congestion effect does not occur on the same DMUs. Therefore, we draw the following policy implications in the event that output direction is decided in advance, ($\delta_1 = \delta_2 = \delta_3 = \delta_4 = 1$).

For DMU_4, decreasing directional RTS prevails when staff and funds increase with the proportion located in the R1 region in Figure 18.2, which means its inputs should not be expanded any further such that the efficiency of resource utilisation of DMU_4 can be further improved.

We also note that the directional congestion effect exists on the right-hand side of DMU_{33} in certain directions, as shown in Table 18.7, which means if the inputs of DMU_{33} are further expanded alongside these directions, the outputs will decrease. Therefore, we conclude that there should not be any additional inputs with these specific proportion of DMU_{33}.

4. CONCLUSIONS AND DISCUSSION

This chapter aims to investigate the directional RTS of the S&T activities of 38 Project 985 universities in China. First, the input–output indicators are proposed, including (1) input indicators such as teaching and research staff (STAFF) and S&T funds (FUND) and (2) output indicators such as monographs, papers, technology transfer income (TT INCOME) and awards. Second, this chapter uses the methods proposed by Yang and Liu (2017) and Yang et al. (2019) to analyse the directional returns to scale and the effect of directional congestion of these universities in China. Based on the analytical results, we detect the region of increasing (constant, decreasing) directional returns to scale for the S&T activities of each university. This information can be used as one of the bases for decision making regarding organisational adjustment. Furthermore, we find that the effect of congestion and directional congestion is evident in the S&T activities of several Project 985 universities. The results of the analysis indicate that directional RTS analysis can provide more detailed information than classic RTS analysis. In other words, the outputs of these universities will decrease as the inputs increase. Therefore, these universities should analyse the underlying reason for the occurrence of the congestion effect, so S&T resources can be used more effectively.

5. ACKNOWLEDGEMENTS

We acknowledge the support of the National Science Foundation of China (71671181, 71201158).

NOTES

1. $P_{BCC} = \{(X, Y) \mid \Sigma_{j=1}^{n} \lambda_j X_j \le X, \Sigma_{j=1}^{n} \lambda_j Y_j \ge Y, \lambda_j \ge 0, \Sigma_{j=1}^{n} \lambda_j = 1\}.$
2. $P_{Convex} = (X, Y) = \{(X, Y) \mid \Sigma_{j=1}^{n} \lambda_j X_j = X, \Sigma_{j=1}^{n} \lambda_j Y_j \ge Y, \lambda_j \ge 0, \Sigma_{j=1}^{n} \lambda_j = 1\}.$

REFERENCES

Atici, K.B., & Podinovski, V.V. (2012). Mixed partial elasticities in constant returns-to-scale production technologies. *European Journal of Operational Research*, 220, 262–269.

Banker, R.D. (1984). Estimating the most productive scale size using data envelopment analysis. *European Journal of Operational Research*, 17, 35–44.

Banker, R.D., Charnes, A., & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092.

Banker, R.D., Chang, H., & Cooper, W.W. (1996a). Equivalence and implementation of alternative methods for determining returns to scale in data envelopment analysis. *European Journal of Operational Research*, 89, 473–481.

Banker, R.D., Bardhan, I., & Cooper, W.W. (1996b). A note on returns to scale in DEA. *European Journal of Operational Research*, 88(3), 583–585.

Banker, R.D., Cooper, W.W., Seiford, L.M., Thrall, R.M., & Zhu, J. (2004). Returns to scale in different DEA models. *European Journal of Operational Research*, 154, 345–362.
Banker, R.D., & Thrall, R.M. (1992). Estimation of returns to scale using data envelopment analysis. *European Journal of Operational Research*, 62(1), 74–84.
Chang, K.P., & Guh, Y.Y. (1991). Linear production functions and the data envelopment analysis. *European Journal of Operational Research*, 52(2), 215–223.
Charnes, A., & Cooper, W.W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, 9, 181–196.
Cooper, W.W., Seiford, L.M., & Zhu, J. (2004). *Handbook on data envelopment analysis*. Massachusetts, USA: Kluwer Academic Publishers.
Cooper, W.W., Huang, Z., & Li, S. (1996). Satisficing DEA models under chance constraints. *Annals of Operations Research*, 66, 279–295.
Cooper, W.W., Gu, B.S., & Li, S.L. (2001). Comparisons and evaluation of alternative approaches to the treatment of congestion in DEA. *European Journal of Operational Research*, 132, 62–67.
Cooper, W.W., Seiford, L.M., & Tone, K. (2000). *Data envelopment analysis*. Boston, USA: Kluwer Academic Publishers.
Färe, R., & Grosskopf, S. (1983). Measuring congestion in production. *Zeitschrilft für Nationalökonomie*, 257–271.
Färe, R., & Grosskopf, S. (1985). A nonparametric cost approach to scale efficiency. *Scandinavian Journal of Economics*, 87, 594–604.
Färe, R., Grosskopf. S., & Lovell, C.A.K. (1983). The structure of technical efficiency. *Scandinavian Journal of Economics*, 85, 181–190.
Färe, R., Grosskopf. S., & Lovell, C.A.K. (1985). *The measurement of efficiency of production*. Boston, USA: Kluwer Nijhoff.
Färe, R., Grosskopf. S., & Lovell, C.A.K. (1994). *Production frontiers*. Cambridge, UK: Cambridge University Press.
Fox, K.J. (2002). *Efficiency in the public sector*. Springer.
Førsund, F.R. (1996). On the calculation of the scale elasticity in DEA models. *Journal of Productivity Analysis*, 7, 283–302.
Fukuyama, H. (2000). Returns to scale and scale elasticity in data envelopment analysis. *European Journal of Operational Research*, 125, 93–112.
Fukuyama, H. (2003). Scale characterizations in a DEA directional technology distance function framework. *European Journal of Operational Research*, 144, 108–127.
Geisler, E. (2000). *The metrics of science and technology*. Quorum Books.
Golany, B., & Yu, G. (1997). Estimating returns to scale in DEA. *European Journal of Operational Research*, 103, 28–37.
Huang, Z., Li, S.X., & Rousseau, J.J. (1997). Determining rates of changes in data envelopment analysis. *Journal of the Operational Research Society*, 48, 591–599.
Kao, C. (2010). Congestion measurement and elimination under the framework of data envelopment analysis. *International Journal of Production Economics*, 123, 257–265.
Keeney, R.L., & Gregory, R.S. (2005). Selecting attributes to measure the achievement of objectives. *Operations Research*, 53(1), 1–11.
Kerstens, K., & Vanden Eeckaut, P. (1998). Estimating returns to scale using nonparametric deterministic technologies: A new method based on goodness-of-fit. *European Journal of Operational Research*, 113, 206–214.
Khodabakhshi, M., Gholami, Y., & Kheirollahi, H. (2010). An additive model approach for estimating returns to scale in imprecise data envelopment analysis. *Applied Mathematical Modelling*, 34, 1247–1257.
Lozano, S., & Gutierrez, E. (2011). Efficiency analysis and target setting of Spanish airport. *Networks and Spatial Economics*, 11(1), 139–157.
Meyer, M.W. (2002). *Rethinking performance measurement: Beyond the balanced scorecard*. Cambridge University Press.
Meng, W., Mingers, J., & Liu, W.B. (2007). Studies on framework for science-technology evaluation using soft system methodology. *Science research management*, 28(2), 1–8. (in Chinese).
Mingers, J., Liu, W.B., & Meng, W. (2009). Using SSM to structure the identification of inputs and outputs in DEA. *Journal of the Operational Research Society*, 60(2), 168–179.
Pindyck, R.S., & Rubinfeld, D.L. (2000). *Microeconomics*. Prentice Hall, 4th Edition.
Podinovski, V.V., & Førsund, F.R. (2010). Differential characteristics of efficient frontiers in data envelopment analysis. *Operations Research*, 58(6), 1743–1754.
Poister, T.H. (2003). *Measuring performance in public and nonprofit organizations*. San Francisco, Calif.: Jossey-Bass.
Read, L.E., & Thanassoulis, E. (2000). Improving the identification of returns to scale in data envelopment analysis. *Journal of the Operational Research Society*, 51, 102–110.
Roper, S., Hewitt-Dundas, N., & Love, J.H. (2004). An ex ante evaluation framework for the regional benefits of publicly supported R&D projects. *Research Policy*, 33, 487–509.
Seiford, L.M., & Zhu, J. (1998). On piecewise loglinear and log efficiency measures. *Computers and Operations Research*, 25(5), 389–395.

Seiford, L.M., & Zhu, J. (1999). An investigation of returns to scale in data envelopment analysis. *Omega-The International Journal of Management Science*, 27, 1–11.
Soleimani-damaneh, M. (2012). On a basic definition of returns to scale. *Operations Research Letters*, 40, 144–147.
Soleimani-damaneh, M., Jahanshahloo, G.R., & Reshadi, M. (2006). On the estimation of returns-to-scale in FDH models. *European Journal of Operational Research*, 174, 1055–1059.
Sueyoshi, T. (1999). DEA duality on returns to scale (RTS) in production and cost analyses: An occurrence of multiple solutions and differences between production-based and cost-based RTS estimates. *Management Science*, 45, 1593–1608.
Sueyoshi, T., & Sekitani, K. (2005). Returns to scale in dynamic DEA. *European Journal of Operational Research*, 161, 536–544.
Sueyoshi T., & Sekitani K. (2007). Measurement of returns to scale using a non-radial DEA model: A range-adjusted measure approach. *European Journal of Operational Research*, 176, 1918–1946.
Tone, K. (1996). A simple characterization of returns to scale in DEA. *Journal of the Operations Research Society of Japan*, 39, 604–613.
Tone, K., & Sahoo, B.K. (2003). Scale, indivisibilities and production function in data envelopment analysis. *International Journal of Production Economics*, 89, 165–195.
Tone, K., & Sahoo, B.K. (2004). Degree of scale economies and congestion: A unified DEA approach. *European Journal of Operational Research*, 158, 755–772.
Wei, Q.L., & Yan, H. (2004). Congestion and returns to scale in data envelopment analysis. *European Journal of Operational Research*, 153, 641–660.
Yang, G.L. (2012). *On relative efficiencies and directional returns to scale for research institutions*. Ph.D thesis, University of Chinese Academy of Sciences, Beijing. (in Chinese).
Yang, G.L., Liu, W.B. (2017). Estimating directional returns to scale in DEA. *INFOR: Information Systems and Operational Research*, 55(3), 243–273.
Yang, G.L., Ren, X.T., Khoveyni, M., & Eslami, R. (2019). Directional congestion in data envelopment analysis. *Journal of Management Science and Engineering*, 5(1), 57–75.
Zarepisheh, M., Khorram, E., & Jahanshahloo, G.R. (2010). Returns to scale in multiplicative models in data envelopment analysis. *Annals of Operations Research*, 173, 195–206.
Zhang, D.Q., Yang, G.L., & Li, X.X. (2011). Strategy maps and performance measures for national research institutes. *Studies in Science of Science*, 29(12), 1835–1844. (in Chinese).
Zhu, J., & Shen, Z.H. (1995). A discussion of testing DMUs' returns to scale. *European Journal of Operational Research*, 81, 590–596.

19. Political short termism and government spending efficiency in Sub-Saharan Africa[1]

Sijuola Orioye Olanubi and Oluwanbepelumi Esther Olanubi

1. INTRODUCTION

Why is there large-scale inefficiency in government expenditure on education and health in Sub-Saharan Africa (SSA)? From a political economy perspective, the answer to this question lies in the high rate of corruption and poor institutional quality in the region. This is also in line with findings in the literature on the determinants of public sector efficiency (see for instance Hauner and Kyobe, 2008; Fonchamnyo and Sama, 2016; Novignon, 2015). This argument, however, weakens when one considers that public goods are both grossly underfunded in Africa and are not major targets for governments' rent-seeking activities as is the case with expenditure on infrastructure, government administration, defence, public works or transfers.[2] Moreover, since improvement in education and health outcomes do not make governments in SSA appear competent before voters whose performance scorecard of an administration is based majorly on visible short-term outcomes, politicians are not under pressure to divert funds to education and health in order to 'pass the performance test' required for re-election.[3] As such, corruption or weak institutions alone cannot explain the high level of inefficiency in government spending on the public goods observed in SSA countries.

We argue in this chapter that inefficiency in government expenditure on education and health is not so much because of corruption, but is due to the neglect of the sectors by politicians because both outcomes do not have a significant impact on their chances of being re-elected. In essence, the inefficiency observed in government expenditure on both outcomes might be due to 'political short termism' or policy myopia, a situation where governments devote a substantial amount of their resources to short-term public goods that generate immediate and visible returns in order to increase their chances of being re-elected, and neglect long-term goods such as education and health that only deliver returns in the future, well after they must have left office. According to Aidt and Dutta (2007), this phenomenon began in the 1930s and became entrenched in society in the 1960s and 1970s, when governments shifted to transfers as a form of redistribution, away from long-term investment in health, education and housing.

Our results support this proposition. We find that corruption is not a significant factor influencing the efficiency of government spending on both education and health while bureaucracy quality has a positive effect on efficiency. This suggests that the root cause of inefficiency in spending on development outcomes observed in the region is not directly a result of corruption but is related to the 'myopic' behaviour of politicians who neglect investment in long-term public goods and divert resources (both in terms of their financial and administrative capacity) to short-term outcomes that deliver visible returns. Among other implications, the neglect of these outcomes will result in poor financial management practices among bureaucrats in the sectors and also serve as a disincentive for labour to perform their activities effectively. Olanubi and Osode (2017) for instance found large-scale inefficiency in government expenditure on human resources for health in Nigeria.

The rest of this chapter proceeds as follows. In section 2, we present a brief review of the literature on political short termism and that on the determinants of government spending efficiency. Section 3 presents the methodology and model specification of the study. Section 4 presents and discusses the results of our analysis. In section 5, we conclude the study and make important policy recommendations.

2. BRIEF REVIEW OF RELEVANT LITERATURE

This study is connected to two related strands of literature that have hitherto developed along separate lines of inquiry. The first is the literature on political short termism or policy myopia and the second is the literature on the determinants of public sector efficiency. When fused together, the two areas of literature explain the reason for the inefficiency in government spending on development outcomes, specifically on education and health, observed in SSA countries.

The literature on political short termism dates as far back as Rogoff (1990), who characterized the phenomenon as occurring when an incumbent, desiring to appear competent before voters, coordinates fiscal policy in a manner that favours short-term consumption expenditures rather than long-term investment. This view is supported by Gersbach (2004), who connects short termism to the low discount rate of politicians. He argues that voters and politicians do not have the same discount rates. As such, voters cannot persuade politicians with a low discount rate to commit to investing in long-term public goods. Also, according to Darby et al. (2004), short termism arises due to political uncertainty as politicians shift the focus of their spending activities towards current consumption and away from investment in long-term public goods that cannot deliver significant returns during their tenure in office.

Short termism has also been associated with the rent-seeking behaviour of corrupt politicians. Mauro (1998) and Garrì (2010) both argue that corrupt politicians have an incentive to divert public resources to expenditure items that make it easier to take bribes without being caught. In particular, Mauro (1998) argues that that expenditure on education is characterized by large-scale corruption while according to Garrì (2010), short termism occurs because investment in long-term public goods makes it easier for politicians to divert resources. This is, however, not the case in SSA, where long-term public goods like education and health are grossly underfunded by the government.

Aidt and Dutta (2007) simplify the assumption advanced by Gersbach (2004) by arguing that both politicians and voters have the same discount rates and that short termism arises due to the interplay of economic and political factors such as observation lags, economic growth, revenue constraint and political uncertainty, and is not due to asymmetries in information, as advanced by previous studies. The study also showed that the diversion of funds by politicians occurs across all forms of public expenditure, and is not related to short- or long-term spending items.

Our study relates to the stream of literature advanced by Rogoff (1990), Gersbach (2004) and Darby et al. (2004) as it focuses on political short termism that arises because politicians prioritize spending on short-term outcomes that deliver immediate returns as against long-term public goods, which is the experience of countries in SSA.

The literature on the determinants of public sector efficiency has also evolved along several areas of emphasis. Rayp and Van De Sijpe (2007) estimated the efficiency of government spending in 52 developing countries and also went ahead to examine its determinants. Their results reveal that structural country variables and governance indicators are the major determinants of government spending efficiency in developing countries. In particular, they find that the adult illiteracy rate and a large population of young people in the population are associated with low levels of spending efficiency while urbanization has a positive effect on efficiency. Good governance which is connected to the rule of law and political stability also has a positive effect on efficiency.

Adam et al. (2011) examined the determinants of efficiency in 19 OECD countries and found that the quality of governance is the major determinant of efficiency, ahead of socioeconomic environment factors. Feeny and Rogers (2008) found that governance and literacy are the major determinants of government spending efficiency. Hauner and Kyobe (2008) examined the determinants of government spending efficiency on education and health in a panel of 114 countries from 1980 to 2006. The determinants were categorized into three areas: economic, institutional, and demographic and geographic. Their results reveal the following. First, that higher government expenditure has a negative effect on efficiency. Second, that strong institutions affect efficiency through economic growth and financial development. Specifically, they find that government accountability and low corruption have a positive effect on efficiency. They also find that demographic and geographic factors have a significant effect on efficiency.

More recently, Fonchamnyo and Sama (2016) assessed the efficiency of government expenditure on education and health in three African countries: Cameroon, Chad and Central African Republic. They also examined the economic and institutional factors affecting the efficiency levels in both sectors. The Data Envelopment Analysis (DEA) technique was adopted in the first stage to generate efficiency scores while the determinants of efficiency were analysed in a second stage using the panel data tobit and the Fractional Logit regression techniques. Their results reveal that corruption has a negative and statistically significant effect on public spending efficiency while the quality of budgetary and financial management practices leads to improvement in spending efficiency. For SSA, Novignon (2015) examined the effects of corruption and quality institutions on the efficiency of public funds allocated to health. A two-stage analysis was adopted, with efficiency scores generated in the first stage from the DEA while the determinants of efficiency were analysed using a tobit model. Results of the study reveal that corruption has a negative effect on efficiency while institution has a positive effect. This study, however, only focuses on the political economy related determinants of government spending efficiency; particularly, corruption and bureaucracy quality, as they are the determinants related to the literature on political short termism.

3. METHODOLOGY AND MODEL SPECIFICATION

Government spending efficiency can be estimated through two main techniques: Stochastic Frontier Analysis and Data Envelopment Analysis (DEA). This study adopts Stochastic Frontier Analysis (SFA). As an econometric approach, SFA has the advantage of being able to separate random noise from inefficiency, unlike the non-parametric DEA, that is unable to do so.

We also adopt the model by Battese and Coelli (1995) that provides a single-step approach to estimating both technical efficiency and its determinants as against the two-step approach where efficiency estimates are first of all generated before determinants are estimated. As demonstrated by Wang and Peter Schmidt (2002), the two-step approach is biased, with a high degree of severity. The one-step approach is therefore strongly advocated.

The Inefficiency Frontier Model for Panel Data by Battese and Coelli (1995) is thus specified:

Consider the stochastic frontier production function:

$$Y_{it} = \exp\left(X_{it}\beta + V_{it} - U_{it}\right) \tag{19.1}$$

Where Y_{it} represents production at the t-th observation (t = 1,2,. . . . .T) for the i-th firm (i = 1,2. . . . .N), X_{it} is a (1 × K) vector for the values of input variables as well as other explanatory variables related to the i-th firm at the t-th observation.

β is a (K × 1) vector of unknown parameters to be estimated.

The V_{it} values are assumed to be iid (0, σ_v^2) random errors which are independently distributed of the U_{it} values.

The U_{it} values are non-negative random variables associated with technical inefficiency of production, and are assumed to be independently distributed.

Z_{it} is a (1 × m) vector of explanatory variables associated with technical inefficiency of production of firms over time.

Lastly, δ is an (m × 1) vector of unknown coefficients.

The technical inefficiency effects, the U_{it} values, are assumed to be a function of a set of explanatory variables, the Z_{it} values and an unknown vector of coefficient δ.

The technical inefficiency effect U_{it} in the stochastic frontier model (19.1) can be specified as

$$U_{it} = Z_{it}\delta + W_{it} \tag{19.2}$$

where the random variable W_{it} is defined by the truncation of the normal distribution with zero mean and variance σ^2.

The likelihood function is expressed in terms of variance parameter $\sigma_S^2 \equiv \sigma_v^2 + \sigma^2$ and $\gamma \equiv \sigma^2/\sigma_S^2$.

The technical efficiency of production of the i-th firm at the t-th observation is therefore defined as

$$TE_{it} = \exp\left(-U_{it}\right) = \exp\left(-Z_{it}\delta - W_{it}\right) \tag{19.3}$$

3.1 Model Specification

The stochastic frontier production function to be estimated is specified below in line with Battese and Coelli (1995).

$$PSP_{it} = \beta_0 + \beta_1\left(GEDUC_{it}\right) + \beta_2\left(GHEALTH\right) + \beta_3\left(YEAR\right) + V_{it} - U_{it} \tag{19.4}$$

The technical inefficiency effects are assumed to be:

$$U_{it} = \delta_0 + \delta_1 \left(CRRP_{it} \right) + \delta_2 \left(\text{BURCY} \right) + \delta_3 \left(YEAR_{it} \right) + W_{it} \tag{19.5}$$

where;
PSP is the public sector performance variable for education and health
GEDUC is government spending on education
GHEALTH is government spending on health
CRRP is corruption.
BURCY is bureaucracy quality
Year is the year of the observation involved
and V_{it} and U_{it} are as defined earlier.

In line with Afonso, Schuknecht and Tanzi (2005) and Afonso, Jalles and Venâncio (2019) we construct a composite for **Public Sector Performance** (PSP) which captures government spending on two major long-term public goods: education and health. The indicator therefore includes two main sub-indicators: the sub-indicator on education and the sub-indicator on health. For the **health** sub-indicator, data for infant survival rate and life expectancy was adopted while the **education** sub-indicator was proxied by data on secondary school enrolment rate due to data availability. Also, a five-year average was computed for the two sub-indicators in order to account for structural change.[4]

Corruption and bureaucracy quality were included in our model as our study only focuses on the political economy-related determinants of government spending efficiency. Also, in line with Battese and Coelli (1995), we include a year variable in both the stochastic frontier and the inefficiency effects section. This is done in order to account for both technical change and time-varying inefficiency effects.

3.2 Data and Variables

The dataset adopted in this study includes 29 countries in SSA. The choice of countries is solely based on data availability. The time frame covered by the study spans four distinct periods: 1998–2002, 2003–07, 2008–12, 2013–17. See Appendix A for detailed information on the sources of data and the description of the variables adopted in the study.

4. EMPIRICAL RESULTS

4.1 Maximum Likelihood Estimates of the Technical Inefficiency Effects Model

Table 19.1 presents results of the maximum likelihood estimates of the technical inefficiency effects model. The first step is to test for the presence of technical inefficiency in the model by comparing the likelihood ratio of the technical inefficiency model against that of the Generalized Linear Model (GLM). The test statistic from the likelihood ratio test is 18.48. Since this is greater than the critical value of 5.412 from Kodde and Palm (1986) we reject the null hypothesis of the absence of technical inefficiency effects in the model. This implies the presence of technical inefficiency in the dataset and makes the Battese and Coelli (1995) stochastic frontier model with technical inefficiency appropriate for the analysis, as against the GLM.

As would be expected, government expenditure on education and health has a positive and significant effect on the PSP index for education and health. This implies that an increase in government expenditure on both spending items would result in improved performance in their outcomes. The magnitude of impact is, however, greater for expenditure on health (with a coefficient of 0.04) than for education (with a coefficient of 0.03). This is also supported by results from the GLM as spending on education and health also have a positive effect on the PSP index, with the size of the health coefficient larger than that of education. The year variable is, however, not significant in both models, which implies the absence of technical change in the model.

Table 19.1 Maximum likelihood estimates

Dependent Variable: PSP for Education and Health		
	Model 1	Model 2
Constant	51.06	0.09
	(31.42)	(7.63)
GEDUC	**0.03***	**0.04***
	(0.01)	**(0.01)**
GHEALTH	**0.04****	**0.05***
	(0.02)	**(0.02)**
Year	–0.025	0.0003
	(0.02)	(0.004)
μ		
Constant	**73.05****	
	(30.509)	
CRRP	–0.07	
	(0.04)	
BURCY	**–0.11***	
	(0.04)	
Year	**–0.04****	
	(0.02)	
δ_u^2	**0.12****	
	(0.06)	
δ_v^2	**0.16***	
	(0.03)	
Λ	**0.73***	
	(0.09)	
γ	**0.35**	
Log likelihood	**26.56**	**17.32**

Notes: Model 1 is the stochastic frontier inefficiency model while Model 2 is the Generalized Linear Model (GLM). Standard errors are reported in parenthesis. *Statistically significant at the 1 per cent level. **Statistically significant at the 5 per cent level. ***Statistically significant at the 10 per cent level.

We move to the technical inefficiency section where we estimate the effects of two political economy related variables (corruption and bureaucracy quality) on efficiency. The results show that corruption has no significant effect on the efficiency of government spending on both education and health while bureaucracy quality has a positive effect on efficiency (or a negative effect on inefficiency).[5] This suggests that the root cause of inefficiency in spending on both development outcomes is not directly related to corruption but to political short termism that makes government neglect investment in long-term public goods and divert resources to short-term outcomes that deliver visible returns. It is, however, important to note that, while the relationship between bureaucracy quality and efficiency is consistent with findings in the literature, that between corruption and efficiency is not. For instance, in contrast to our results, studies in the literature conducted for Africa, such as Fonchamnyo and Sama (2016) and Novignon (2015), find a significant negative effect of corruption on efficiency. The reason for this is far fetched. First, both studies adopted the two-step analysis in estimating the determinants of efficiency which has been adjudged to generate biased estimates. Also, the sample size and the choice of variables adopted in this study differs from theirs. Fonchamnyo and Sama (2016) examined the determinants of government spending efficiency on education and health in three African countries: Cameroon, Chad and Central African Republic while Novignon (2015), who explored a large sample of 45 countries in SSA, only focused on the determinants of government spending efficiency on health.

The year variable has a negative sign and is significant, showing that inefficiency reduces over time, which is the same as reporting that efficiency increases over time in the countries considered. This is supported by findings in section 4.2 where efficiency estimates for all countries increased over the four five-year periods.

The estimate of γ is 0.35. This suggests that 35 per cent of variation in the composite error term is accounted for by the inefficiency component. The relatively low value is due to the absence of control variables in the model. The inclusion of more control variables aside from those directly related to government spending on education and health would bias our estimates since we are particularly interested in the effect of government expenditure on the two items.

4.2 Efficiency Estimates for the Education and Health PSP Index

Table 19.2 presents efficiency estimates for the 29 SSA countries considered in this study. Average efficiency for all countries is 0.794. This implies that, on average, countries in SSA have wasted 20.6 per cent of public funds allocated to education and health during the study period. Also, Niger had the least efficiency score, at 0.69, while South Africa has the highest efficiency score, at 0.93. In other words, while South Africa has been able to efficiently utilize 93 per cent of public funds allocated to the outcomes Niger has only been able to efficiently utilize 69 per cent of the funds. This is not surprising as South Africa has a well-developed political system, strong institutional arrangements and efficient productive capacities in comparison to other countries in SSA (such as Niger) characterized by political instability, weak institutions and inefficient productive capacities. Also, efficiency estimates for all countries increased over the years, in line with findings of the stochastic frontier of the positive effect of the year variable on efficiency. An explanation for this is the lag between when expenditures are made on education and health and when they translate into improvements in outcomes.

Table 19.2 Efficiency estimates for selected SSA Countries

Country	Year	Efficiency Score	Country	Year	Efficiency Score
Angola	1998–2002	0.558109	Madagascar	2008–12	0.877212
Angola	2003–07	0.727005	Madagascar	2013–17	0.928343
Angola	2008–12	0.816794	Malawi	1998–2002	0.672121
Angola	2013–17		Malawi	2003–07	0.747611
Botswana	1998–2002		Malawi	2008–12	0.878424
Botswana	2003–07	0.889976	Malawi	2013–17	0.944331
Botswana	2008–12	0.934802	Mali	1998–2002	0.536604
Botswana	2013–17		Mali	2003–07	0.645864
Burkina Faso	1998–2002		Mali	2008–12	0.781854
Burkina Faso	2003–07	0.647669	Mali	2013–17	0.890267
Burkina Faso	2008–12	0.782282	Mozambique	1998–2002	0.515353
Burkina Faso	2013–17	0.911287	Mozambique	2003–07	0.621807
Cameroon	1998–2002	0.628936	Mozambique	2008–12	0.771391
Cameroon	2003–07	0.724596	Mozambique	2013–17	0.895059
Cameroon	2008–12	0.905715	Namibia	1998–2002	0.728492
Cameroon	2013–17	0.952845	Namibia	2003–07	0.79105
Democratic Republic of Congo	1998–2002		Namibia	2008–12	0.857476
Democratic Republic of Congo	2003–07		Namibia	2013–17	0.949677
Democratic Republic of Congo	2008–12	0.798117	Niger	1998–2002	0.51036
Democratic Republic of Congo	2013–17	0.900907	Niger	2003–07	0.611472
Cote d'Ivoire	1998–2002	0.598616	Niger	2008–12	0.761324
Cote d'Ivoire	2003–07	0.653308	Niger	2013–17	0.883058
Cote d'Ivoire	2008–12	0.769889	Senegal	1998–2002	0.608593
Cote d'Ivoire	2013–17	0.879407	Senegal	2003–07	0.70094
Ethiopia	1998–2002	0.566765	Senegal	2008–12	0.831723
Ethiopia	2003–07	0.683104	Senegal	2013–17	0.929581
Ethiopia	2008–12	0.853049	Sierra Leone	1998–2002	0.545584
Ethiopia	2013–17	0.930175	Sierra Leone	2003–07	0.635119

(*Continued*)

Table 19.2 (Continued)

Country	Year	Efficiency Score	Country	Year	Efficiency Score
Gabon	1998–2002	0.741374	Sierra Leone	2008–12	0.777365
Gabon	2003–07		Sierra Leone	2013–17	0.877412
Gabon	2008–12	0.889987	South Africa	1998–2002	0.897014
Gabon	2013–17	0.94772	South Africa	2003–07	0.922912
Gambia	1998–2002	0.749171	South Africa	2008–12	0.948653
Gambia	2003–07	0.830997	South Africa	2013–17	0.970098
Gambia	2008–12	0.924112	Tanzania	1998–2002	0.634059
Gambia	2013–17	0.946414	Tanzania	2003–07	0.721018
Ghana	1998–2002	0.702027	Tanzania	2008–12	0.830275
Ghana	2003–07	0.794007	Tanzania	2013–17	0.907296
Ghana	2008–12	0.909745	Togo	1998–2002	0.597037
Ghana	2013–17	0.966144	Togo	2003–07	0.697073
Guinea	1998–2002	0.665998	Togo	2008–12	0.818886
Guinea	2003–07	0.790306	Togo	2013–17	0.914693
Guinea	2008–12	0.88976	Uganda	1998–2002	0.6561
Guinea	2013–17	0.943529	Uganda	2003–07	0.825461
Guinea-Bissau	1998–2002	0.557606	Uganda	2008–12	0.898371
Guinea-Bissau	2003–07		Uganda	2013–17	0.94603
Guinea-Bissau	2008–12	0.873437	Congo Republic	1998–2002	0.711003
Guinea-Bissau	2013–17	0.926292	Congo Republic	2003–07	0.77357
Kenya	1998–2002	0.685319	Congo Republic	2008–12	0.862478
Kenya	2003–07	0.782664	Congo Republic	2013–17	0.925402
Kenya	2008–12	0.902696	Zambia	1998–2002	0.618926
Kenya	2013–17	0.945635	Zambia	2003–07	0.742238
Liberia	1998–2002		Zambia	2008–12	0.894905
Liberia	2003–07		Zambia	2013–17	
Liberia	2008–12	0.828456	Zimbabwe	1998–2002	
Liberia	2013–17	0.898397	Zimbabwe	2003–07	
Madagascar	1998–2002	0.712327	Zimbabwe	2008–12	0.806965
Madagascar	2003–07	0.766302	Zimbabwe	2013–17	0.906149

5. CONCLUSION AND POLICY RECOMMENDATIONS

In this chapter, we examined the relationship between policy myopia and government spending efficiency in a sample of 29 countries in Sub-Saharan Africa. Our argument was founded on the premise that inefficiency in government spending on health and education is not so much because of corruption but due to the neglect of the sectors by politicians because both outcomes do not have any significant impact on their chances of being re-elected. This is supported by our results. We find that corruption is not a significant factor influencing the efficiency of government spending on education and health while bureaucracy quality has a positive effect on efficiency. This suggests that the root cause of inefficiency in spending on development outcomes observed in the region is not directly related to corruption but to the 'myopic' behaviour of politicians who neglect investment in long-term public goods and divert resources to short-term outcomes that deliver visible returns. Among other implications, the neglect of these outcomes will result in poor financial management practices among bureaucrats in the sector and also serve as a disincentive for labour to perform their activities effectively. It is therefore recommended that, while countries should continue to clamp down on corruption in the public sector and improve the quality of their bureaucracy, dealing with the problem of policy myopia is crucial to ensuring improvements in the efficiency of government spending on long-term public goods, particularly on education and health. This will involve pursuing policies that make political positions unattractive for rent-seeking activities by reducing remuneration of political offices and cutting down on the cost of governance. This will ensure that only politicians that prioritize the interest of their nation, demonstrated by their willingness to invest in long-term public goods, will be those that will be elected.

APPENDIX A: VARIABLE SOURCES AND DESCRIPTION

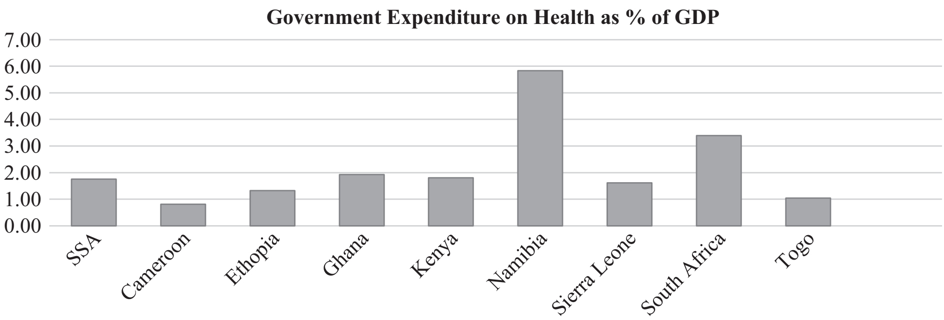
Appendix 19.A1 Output variables

S/N	Sub Index	Variable	Source	Description
1.	External Variables	Corruption	International Country Risk Guide (ICGR) (1998−2017)	Five-year average of corruption Index ranging from 0 to 6. The higher the index, the less corrupt the political system is.
		Bureaucracy Quality	International Country Risk Guide (ICGR) (1998−2017)	Five-year average of Bureaucracy Quality Index ranging from 0 to 4. The higher the index, the stronger the quality of bureaucracy in a country.
2.	Education	Secondary School Enrolment	World Bank, World Development Indicators (1998−2017)	Five-year average of ratio of total enrolment in secondary education.
3	Health	Infant survival rate	World Bank, World Development Indicators (1998−2017)	Five-year average of infant survival rate Infant survival rate = (1000−IMR)/1000. Note that IMR is the infant mortality rate measured per 1000 lives birth in a given year.
		Life Expectancy	World Bank, World Development Indicators (1998−2017)	Five-year average of life expectancy at birth.

Appendix 19.A2 Input variables

S/N	Sub Index	Variables	Source	Description
2.	Education	Government Education Expenditure	World Bank, World Development Indicators (1998−2017)	Five-year average of expenditure on education (% of GDP)
3.	Health	Government Health Expenditure	World Bank, World Development Indicators (1998−2017)	Five-year average of expenditure on health (% of GDP)

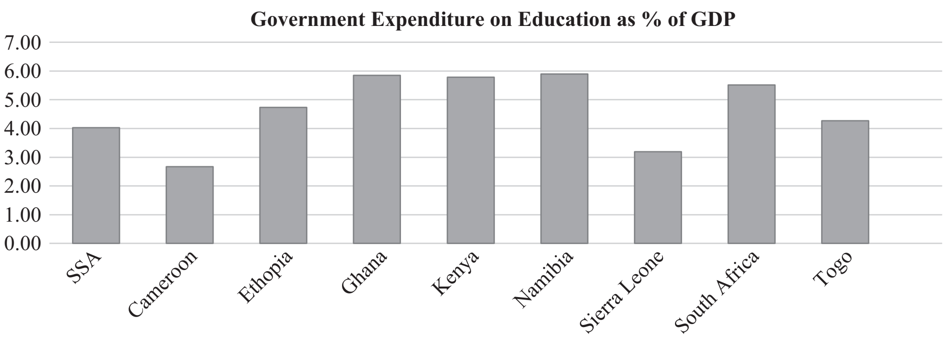
APPENDIX B: STYLIZED FACTS ON HEALTH AND EDUCATION EXPENDITURE IN SSA



Note: Average amount expended on health by governments in Sub-Saharan African countries (in the 29 countries considered in this study) as a share of their GDP between 1998 and 2017 is below 2 per cent (at 1.76 per cent).

Source: World Bank's World Development Indicators.

Figure 19.B1 Government expenditure on health in SSA (1998−2017)



Note: Average amount expended on education by governments in Sub-Saharan African countries (in the 29 countries considered in this study) as a share of their GDP between 1998 and 2017 is 4 per cent. `

Source: World Bank's World Development Indicators.

Figure 19.B2 Government expenditure on education in SSA (1998−2017)

NOTES

1. The views expressed in this chapter are those of the authors and do not represent those of the Nigerian Economic Summit Group or its Board of Directors.
2. Figures B1 and B2 in Appendix B reveal that the average amount expended on health by governments in Sub-Saharan African countries as a share of their GDP between 1998 and 2018 is below 2 per cent (at 1.76 per cent) while the figure for education is 4 per cent.
3. The diversion of funds by the government to a spending item creates an avenue for the misappropriation of funds by bureaucrats and government officials.
4. See Afonso, Schuknecht and Tanzi (2005) and Afonso, Jalles and Venâncio (2019) for detailed explanations on the construction of the PSP.
5. Note that, in the inefficiency effects section of the results, the effect of the determinants is interpreted as affecting inefficiency and not efficiency, hence the negative sign of the coefficient of the variables. But we have made our interpretation to focus on efficiency.

REFERENCES

Adam, A., Delis, M., & Kammas. (2011). Public sector efficiency: Leveling the playing field between OECD countries. *Public Choice*, *146*, 163–183.
Afonso, A., Jalles, J., & Venâncio, A. (2019). Taxation and public spending efficiency: An international comparison. *REM Working Paper*, 080-2019.
Afonso, A., Schuknecht, L., & Tanzi, V. (2005). Public sector efficiency: An international comparison. *Public Choice*, *123*, 321–347.
Aidt, T.S., & Dutta, J. (2007). Policy myopia and economic growth. *European Journal of Political Economy*, *23*(3), 734–753.
Battese, E., & Coelli, T. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, *20*, 325–332.
Darby, J., Li, C.-W., & Muscatelli, A.V. (2004). Political uncertainty, public expenditure and growth. *European Journal of Political Economy*, *20(1)*, 153–179.
Fonchamnyo, D. C., & Sama, M. C. (2016). Determinants of public spending efficiency in education and health: evidence from selected CEMAC countries, *Journal of Economics and Finance*, 40, 199–210.
Feeny, S., & Rogers, M. (2008). 'Public sector efficiency, foreign aid and small Island developing states.' *Journal of International Development*, *20*(4), 526–546.
Garrì, I. (2010). Political short-termism: A possible explanation. *Public Choice*, *145*, 197–211.
Gersbach, H. (2004). Competition of politicians for incentive contracts and elections. *Public Choice*, *121*, 157–177.
Hauner D., & Kyobe A. (2008). Determinants of government efficiency. *IMF Work Paper*, WP/08/228.
Kodde, D.A., & Palm, F.C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*, *54*, 1243–1248.
Mauro, P. (1998). Corruption and the composition of government expenditure. *Journal of Public Economics*, *69*(2), 263–279.
Novignon, J. (2015). On the efficiency of public health expenditure in Sub-Saharan Africa: Does corruption and quality of public institutions matter? MPRA Paper No. 391–95.
Olanubi, S.O., & Osode, O.E. (2017). The efficiency of government spending on health: A comparison of different administrations in Nigeria. *Journal of Policy Modeling*, *39*(1), 79–98.
Rogoff, K. (1990). Equilibrium political budget cycles. *American Economic Review*, *80(1)*, 21–36.
Rayp, G., & Sijpe, N.V.D. (2007). Measuring and Explaining Government Efficiency in Developing Countries.' *Journal of Development Studies*, *43(2)*, 360–381.
Wang, H., & Schmidt, P. (2002). One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis*, *18(2)*, 129–144.

Index