Universitat Autònoma de Barcelona

Facultat d'Economia i Empresa UAB

Treball de Final de Grau

Facultat d'Economia i Empresa

TÍTOL: Tax effort stochastic frontier estimations: determinants, uses, and issues.

AUTOR/A: *Ivan Yepifanov*

TUTOR/A: *Luca Salvadori*

GRAU: Bachelor's degree in Economics

DATA: *May 2024*

Abstract

Estimating the tax effort – that is the ratio between the actual tax revenues collected by a country and the correspondent potential tax revenues – is extremely important, particularly in the context of the developing world but more generally for internal policy decisions. However, this exercise entails several challenges. Some studies have recently introduced the use of stochastic frontier analysis to estimate the tax effort and a recent study by the UNU-WIDER has revisited the issue and provided new estimates for the largest sample of countries and years up to this date. Building on this strand of the literature, this paper aims to identify additional potential variables not directly employed in this estimation and it proposes refined estimates of the tax effort focusing in particular on the level of development of a country. Indeed, the analysis suggests that some variables seem to correlate differently with the tax effort based on the level of development. Among the candidates, two main variables are found to play a significant role in the estimation of the tax collection frontier and tax effort. These are foreign direct investments (FDIs) and the share of the shadow economy.

Keywords: tax effort, revenue mobilization, potential tax revenue, tax frontiers.

Table of Contents

List of Tables

List of Figures

1. Introduction

In the context of modern economy and innovation, new opportunities to modernize tax collection and tax agencies have been arising continuously. Nevertheless, even now revenue mobilization is a major concern for developing countries. Despite high national and international efforts tax collection remains at relatively low levels when compared to developed nations as demonstrated in *Figure 1*.

Figure 1: Tax collection by income group.

Source: UNU-WIDER (2023).

It is important to highlight that in this case, tax collection is not the total government revenue but the revenue-to-income ratio. Usually, the economic income is measured as gross domestic product (GDP). The tax collection ratio is crucial for countries since it conditions the degree of policy freedom and achievement of sustainable development goals. For example, investments in human capital, infrastructure, provision of public goods and services as well as objectives of energy transition and environmental sustainability are simply unreachable for most low-income and lower middle-income countries. As of 2023, 86% of low-income and 43% of lower-middle-income countries are below the sustainable tax-to-GDP threshold of 15%. This is commonly considered a tipping point of the tax ratio to achieve sustainable development (Gaspar et al., 2016; World Bank, n.d.-b). Therefore, the question of how developing countries can achieve the desired levels of taxation from both equity and efficiency perspectives has been studied extensively.

This paper focuses on the efficiency side of tax systems, in particular on the estimation of the theoretically possible maximum tax revenues. One of the key ideas underlying maximum tax collection is the concept of the Laffer curve (see e.g. Hemming & Kay, 1980) which imputes that for any given country there exists a combination of tax rates that will generate the highest total tax revenue. This theory does not consider equity considerations and focuses only on the efficiency side or in this case maximization of government revenues. In practice, tax agencies are not concerned only with maximizing collections and are influenced by political constraints e.g. the maximum tax rate for subsistence farmers. Nevertheless, conflicting objectives and constraints do not make the study of theoretical maximum tax collection irrelevant. In the quest for development, economists must understand the relative rankings of countries on the tax burden scale and their potential (maximum) tax collections. Potential tax collection is especially useful as it allows governments to measure the gap between current and potential revenues as well as find ways to close it and reach a desirable level. Estimating the potential revenue and searching for ways to mobilize it is the focus of organizations such as the Addis Tax Initiative, which aims to close gaps in development finance through improvements in tax systems.

The common measure of tax revenue inefficiency is called tax effort score which is constructed by dividing observed tax revenues by econometrically estimated the potential tax revenues. The ways to estimate the potential have been evolving continuously and will be reviewed later in this paper. At this point, the state-of-the-art method of estimation is Stochastic Frontier Analysis (SFA) which uses a production function with underlying economic variables to estimate the inefficiency term and then potential tax revenues. The inefficiency score given by this analysis provides a tax effort estimate between 0 and 1, where values close to 1 indicate that the country's actual tax collection is close to the theoretically maximum possible value. The variables used in the frontier estimations are macroeconomic, demographic, and institutional factors. This combination allows researchers to provide international estimations to be used in benchmarking and the process of revenue mobilization. However, the scores vary significantly depending on the method of estimation and the variables included.

One of the most recent studies on this topic is provided by McNabb et al. (2021) as part of the United Nations University World Institute for Development Economics Research and is based on the latest Government Revenue Dataset to estimate the tax effort scores for 192 countries and the period from 1980-2019. The paper provides 4 different tax effort scores, one for each method of estimation employed for the inefficiency parameter and stochastic shocks. By using these new estimates, particularly those obtained from the True Random Effects frontier as they are argued to be superior, this TFG aims at extending the analysis by McNabb et al. (2021) to first look at the apparent correlations between reported tax effort scores and variables not included in their estimation but likely to be relevant and then replicate the stochastic frontier design to estimate tax effort scores including statistically significant omitted variables. In this way, I aim to contribute to the field of tax effort estimations by first testing the robustness of the estimates, outlining potential issues, and proposing an improved version of those estimates.

The rest of the paper will be structured as follows. In section 2 I will provide the literature review on the creation of tax potential estimations. Section 3 will explain the estimation strategy used to achieve the objectives set in this paper. Next, in section 4 data sources and general descriptive statistics will be presented. In section 5, empirical analysis will be conducted, starting with graphical patterns, and linear regression for the general sample and sub-samples per income group. In section 6 the stochastic frontier will be estimated, and tax efforts will be provided following a discussion about the importance of omitted variables. Lastly, in section 7 the paper will be concluded.

2. Literature review

To the best of my knowledge, the literature focused on the explanation of the tax effort in the context of a multi-country framework starts with Ohshima (1957) and ends with McNabb et al. (2021). The latest paper is the one which provides the tax effort scores used in estimations. Moreover, recent studies focused on the stochastic frontier approach and not on regression analysis are Fenochietto & Pessino (2013); Langford & Ohlenburg (2016); Mawejje & Sebudde, (2019). This framework is a significant development although as argued by McNabb et al. (2021) it is far from perfect as there appear to be important flaws and an overestimation of tax potential.

Before looking at the history of the literature it is important to understand how the variables are selected for estimations. The tax effort refers to the maximum potential tax collection given underlying conditions. This implies that researchers focus on identifying those structural items. Therefore, when estimating this theoretical potential, variables are only relevant if they have explanatory power for the tax-to-GDP ratio or any other ratio depending on the metric of

economic well-being used. A variable capable of explaining the aggregate tax collection of a country such as population might not be important in the context of this study as there is a straightforward relationship in nominal terms between taxes collected and the size of the population. However, the population will not have an impact on relative taxes collected by the government as it is not a structural variable. Thus, in this section, only variables proven to affect the relative taxation are considered. Apart from outlining the history of the literature throughout this section, I will identify the variables to be used in my estimations. The variables are going to be selected by examining various studies that made contributions by relating government revenues to economic variables and finding statistical relationships.

Commencing with Oshima (1957) the determinants of tax effort began to be explored. The study was composed of 32 countries and years between 1948 to 1954. This paper focused on the development level to explain low tax effort scores. Additionally, correlations between tax revenue, the authoritarianism of the government, and the importance of subsistence agriculture were provided. Both were found important in explaining tax collections.

Application of the regression approach to the issue followed. Using a sample of 33 countries Williamson (1961) found a positive relationship between the development of the country (proxied by Gross National Product) and the share of government revenue.

Hinrichs (1965) studied a sample of 60 countries between 1957 and 1960 and argued that the GNP per capita is correlated to the revenue ratio (government revenue over GNP) in all the countries individually. Nevertheless, there is no correlation if the sample is divided into two subsamples (developed and less developed). Finally, the paper finds that for the less developed subset, the measure of openness to trade which in this case is imports over the GNP is a better index to explain the share of the tax revenue. In conclusion, Hinrichs (1965) found that the explanatory variable of the share of tax revenue depends on the development level. GNP per capita is the relevant variable for already developed, and the degree of openness for developing countries.

Papers discussed previously, however, were not comparing the potential revenue effort and the actual revenue, in this sense Lotz & Morss (1970) first aimed to model the revenue ratio and then compare it to the actual tax collection. The first finding is that the share of net exports in GNP or (X-M)/GNP is a better explanatory variable for the level of revenue than other measures of openness or trade considered before. Next, the dependence on natural resource revenues as measured by the share of mineral and petroleum exports to total exports is found not to affect the tax ratio. Additionally, they proved the existence of a positive correlation between tax revenue effort and the share of the revenue collected by non-central authority, thus showing that fiscal decentralization is important for resource collection because of trust in the authority. Finally, they show a significant difference between the actual taxes collected and those predicted by the model, which is positive for roughly half of the countries in the dataset and negative for others. They also note that there are different factors explaining tax ratios. For developed countries, the demand for government services is significant. While for developing countries the relevant factors are those influencing the tax base.

Bahl (1971) contributed significantly to the issue by decomposing the variance of tax ratios into parts explained by tax capacity and tax effort. Using the sample of 49 countries for the years from 1966 to 1968 the paper estimated the regression with significant explanatory variables being the share of agriculture in GDP (negative) and the share of mining exports in total exports (positive) to estimate the tax ratio. As in Lotz & Morss (1970), the paper ranked countries based on the predicted tax capacity and the tax effort, which is the ratio of actual tax to the tax capacity.

During the 1970s the International Monetary Fund provided 3 studies (IMF, 1971, 1975, 1979) to estimate tax capacity. Each report concluded that openness to trade (exports to GDP), economic structure (proxied by share of mineral exports), and income per capita are significant variables to study the tax effort. It was also found that countries experienced significant changes in their tax effort scores between studies, thus demonstrating sample vulnerability.

The first study to employ autoregression of degree 1 was Leuthold (1991) who applied it to panel data of 8 African countries and concluded that it performs better than the simple ordinary least squares model. Additionally, this paper included grants and aid as explanatory variables. Then, Tanzi (1992) using a sample of 83 countries found that the GDP per capita had lost its importance in explaining the tax ratio for the period between 1978 and 1988, while the structure of the economy became more important. Additionally, the paper proves that the share of agriculture in the GDP is a better explanatory variable than GDP per capita and that the debt-to-GDP ratio is also significant. It was found that higher foreign debt is associated with better revenue performance. From a theoretical perspective, this happens because of the debt servicing. Liabilities denominated in foreign currency generate the necessity to accumulate foreign exchange reserves to pay interests. Therefore, the country may choose to reduce imports with the objective of retaining foreign currency but at the same time, it would diminish import taxes. The second way to accumulate those reserves is to impose import tariffs or other trade measures and service debt with the primary budget surplus (Eltony, 2002; Lutz & Tanzi, 1991). Based on this evidence and due to the significance of this variable, the debt-to-GDP ratio will be introduced as an explanatory variable.

Next, Stotsky & WoldeMariam (1997) applied a fixed effects model to a sample of 43 Sub-Saharan African countries between 1990 and 1995 and found that the share of the primary sector in GDP and openness to trade are positively related to tax effort. Additionally, they found that mining share has a negative relationship with the tax effort, which is in opposition to the previous literature.

Incorporation of the role of corruption was first done by Ghura (1998). In an augmented model, incorporating usual variables such as income, openness, mineral dependence, and agricultural share, variables focused on economic policy such as inflation, real exchange rate, structural reform, and human capital were included. By studying a panel of 39 Sub-Saharan African countries between 1985 and 1996, the paper finds that inflation has a significant negative effect, structural reform, and human capital are significantly positively related to the tax ratio. Most importantly, the level of corruption is shown to have a significant negative relationship to the tax ratio. Finally, there is also a negative statistical relationship between being a member of CFA Franc and tax ratio demonstrating the importance of unobserved regional factors. Based on the evidence presented in this paper inflation will be included in the estimations. It is important because inflation is a proxy for the degree of expansion in monetary and fiscal policies. In theory, there are namely four effects the inflation rate has on revenues. The first one is the Tanzi-Olivera effect which states that under inflationary conditions time erodes the real value of taxes collected because they are collected or paid at the end of a predetermined period, while corresponding transactions occur during said period, creating a lag (Olivera, 1967; Tanzi, 1977). The second effect is that fixed excise rates levied on products or other fixed duties may not be properly adjusted with inflation. Thus, real tax collection will be eroded with inflation (Easterly & Schmidt-Hebbel, 1991; Tanzi, 1989). Alternatively, high inflationary pressures induce a re-optimization of asset baskets held by economic agents favouring those that protect from domestic inflation. Therefore, foreign or real assets that tend to escape the usual tax implications are chosen thus reducing the tax base. Lastly, inflation may increase the revenues collected if tax rates are indexed with a significant lag under progressive tax regimes due to shifts of aggregate tax base into brackets with higher tax rates as

they are computed in nominal terms (Alonso & Garcimartín, 2011). Finally, the inflation rate was proven to have a significant negative impact on revenue performance through a combination of these effects.

Another development in the literature was a paper written by Bird et al. (2004), where authors used usual variables that they consider as supply-side factors and added those related to the quality of institutions meant to represent demand-side factors. These factors are corruption, voice, and accountability which were shown to be important to explain revenue performance. The main findings of this paper are that the quality of societal institutions strongly and positively determines the tax revenue effort. Finally, they find evidence of a negative relationship between the shadow economy and the tax effort. From a theoretical perspective, this variable has been studied extensively and has been proven to have a negative effect on the actual tax collection. In 2018 the influence of the shadow economy on tax collection was estimated by Awasthi & Engelschalk (2018). They concluded that a 1% decrease in the shadow economy leads to a 0.125% increase in the tax-to-GDP ratio. This variable is also crucial for analysis because the shadow economy is estimated to be as much as 1/3 of the world's economy. However, there have been many developments since the variable was first proposed in the context of tax effort, specifically it is not used as an explanatory variable in the most recent publications. This is identified as a potential issue as the statistically significant variable is omitted. This can adversely influence the estimations and the results of the SFA models. There is a consensus that the shadow economy is a large source of inefficiency in tax collection and therefore an important component of tax effort (Russell, 2010). Nevertheless, the effect of the shadow economy from the theoretical perspective on the potential tax collection and the frontier estimations is unclear. Therefore, the shadow economy index is another variable to be included in the regression proposed above.

Sen Gupta (2007) extended the sample and applied more robust econometric methods to estimate the tax effort. The application of fixed effects and random effects models on a panel of 105 developing countries for 25 years confirmed those of previous studies. Additionally, foreign aid received was found to be an important and large positive variable in explaining revenue performance. However, institutional variables such as level of corruption, law and order, and political and economic stability are found not to have any significant effects on revenue. The paper also provides evidence that countries relying on direct taxes are performing better in terms of revenue when compared to those relying on indirect taxes such as trade taxes or taxes on goods and services. Moreover, it is shown that results differ if countries are clustered by income level. Additionally, the share of agriculture in GDP in line with previous findings is proven to be negatively related to the tax effort for all income groups, openness to trade is statistically important for low-income countries and middle-income countries, foreign aid has a positive effect only in low-income countries, and corruption or political stability are significant only for low- and middle-income countries. Both Leuthold (1991) and Gupta (2007) included grants or foreign aid in their estimations, moreover, grants are mentioned by McNabb et al. (2021) as an explanatory variable, but the authors decided to exclude it as it was found to not affect the estimations while drastically reducing the number of observations. Nevertheless, to test the statistical significance of the variable in the context of my paper, grants received by the government will be included as an independent variable. Additionally, it was proven that the difference in the sign comes from the effectiveness of uses of the aid or grants, which if used effectively will improve revenue mobilization and if used to finance current consumption will have no effect on revenue performance. Finally, the composition of foreign aid also affects the impact on revenue mobilization. Concessional loans have a positive effect on revenue, while grants are found to reduce revenues (Pivovarsky et al., 2003).

The studies outlined above have all been using what is known as the "traditional method" or in econometrics terms a regression analysis to estimate the tax effort. Pessino & Fenochietto (2010) have employed the first stochastic frontier model to explain the tax effort. This method involves using a production function to model the highest tax revenue a country could obtain. Using this approach the ratio between the maximum and actual revenue can be decomposed into 2 parts. These are the level of inefficiency (explained by factors such as tax policy or legislative weakness) and random error. The paper has estimated the stochastic frontier using 96 countries and 16 years with the traditional set of variables such as GDP per capita, openness, agriculture value added, public education expenditure, income distribution, inflation, and corruption. All these variables were found to be significant and with expected signs as in previous literature. Moreover, this approach allows for the inclusion of determinants of technical inefficiency. Additionally, they found that the level of corruption is positively related to inefficiency.

In 2013 an updated study was made by Fenochietto & Pessino (2013) which included 17 additional countries with natural resources revenue above 25% of total tax revenue for the period from 1991 to 2012. This study has obtained similar results. Additionally, by using the extended sample they observed an increase in the inefficiency parameter.

Next, Cyan et al. (2013) provided a comprehensive assessment of the estimation methods used in previous studies on the tax effort. By using data from 94 countries for the period from 1970 to 2009 they compare the "traditional" regression approach, the stochastic frontier analysis, and the third approach of estimating the tax effort using a budget balance principle. The budget balance principle is obtained by comparing actual tax collections to countries' expenditures which is argued to be a good measure of spending on a desirable public goods level. Authors include measures such as the complexity of the tax system, the measure of tax morale, political fractionalization, and the level of government debt to explain the inefficiency parameter in the stochastic frontier analysis. The traditional model with fixed effects and estimates from stochastic frontier demonstrate a high correlation between each other and a low correlation coefficient to the measure of budget balance. In the same paper, the authors found that capital investment level has a positive relationship with government revenues. This potentially happens due to the expansion of economic activity and subsequently the tax base. Thus, gross fixed capital formation is the next variable that is included in my estimations.

Ángeles Castro & Ramírez Camarillo (2014) applied dynamic panel estimations on a set of 34 OECD countries for the period between 2001 and 2011. Using a traditional set of explanatory variables along economic, structural, institutional, and social dimensions together with the lagged dependent variable they confirm previous findings and argue that tax effort and tax gap are stable over time. Most importantly, they found that the share of the foreign direct investment in gross fixed capital formation has an important negative impact on tax performance. Moreover, foreign direct investment (FDI) was proven to be related to the tax rates. The inflow of FDI is conditioned by the relative corporate tax rates in 'home' and 'receiving' countries and income tax rates (Cassou, 1997). Thus, given this evidence, the FDI measure will be used in the following estimations.

A recent study by Langford & Ohlenburg (2016) employed the SFA approach and first used the data on tax revenues from UNU-WIDER GRD. They used the standard economic variables from the literature as well as many demographic and institutional factors on a panel of 85 countries and the period from 1984 to 2010. The results about economic factors are in line with those expected from previous literature. Additionally, they found statistically significant effects from measures of law and order, corruption, and democratic accountability on the tax ratio.

Mawejje & Sebudde (2019) recently estimated the tax effort from 150 countries and the period from 1996 to 2015 and provided similar estimates to previous studies.

The most recent study of the tax effort is the McNabb et al. (2021) which used the stochastic frontier approach and UNU-WIDER GRD dataset to provide the most complete data of tax effort scores for 161 countries and a total of 3901 observations. They use traditional economic variables which were discussed in previous literature together with demographic and democracy variables and 4 different ways to estimate the frontier which are Pooled, Random Effects (RE), Battese and Coelli (BC), and True Random Effects (TRE). The main conclusion of this paper, however, comes from the application not of simple SFA but using a true random effects approach which is demonstrated to be less influenced by outlying observations, providing more consistent estimation of tax effort indexes and conservative scores (the tax effort is closer to 1) when compared to other methods of estimation. Additionally, the TRE method of estimation allows to account for unobserved heterogeneity of countries and remove it from the inefficiency parameter estimated while it is not accounted for in RE and BC models. This approach drastically changes the results of the estimations and finds scores that are very different compared to previous literature. On average scores are a lot higher and have a much lower variance. These characteristics make this method of estimation superior to previous ones as it allows for a more realistic estimation of the potential tax collections.

Since in this paper, I am using tax effort scores from McNabb et al. (2021), I now present a brief logic behind the selected variables for stochastic frontier estimation. The special attention paid stems from the fact that to replicate the SFA process these variables must be included in order to avoid omitted variable biases.

GDP per capita is a variable meant to measure the level of economic development or sophistication and is expected to be positively correlated to tax effort for two reasons. The first, reason is that according to Wagener's law, the demand for government-supplied services has a high degree of income elasticity. Thus, the sophistication of the economy leads to a higher demand for those services, and in the end, a larger proportion of goods and services is provided by the government. The second reason is that more sophisticated goods and services are easier to tax, and large operations of established companies are much more transparent than those of subsistence farmers.

GDP per capita squared is meant to understand the linearity of the relationship between the GDP and tax collection, the literature has found the negative coefficient associated with this variable suggesting a non-linearity of tax collected with respect to GDP. This means that an increase in income can stimulate an increase in taxes collected but will reach an asymptote at the point where no more services can be provided and when all the economic activity is taxed.

Trade as a percentage of GDP is also an important variable because of well-defined borders import and export taxes are levied without large enforcement problems and in this way the flow of goods across borders provides a large base to be taxed. Additionally, globalization and the World Trade Organisation have facilitated the reduction of tariffs and quotas thus affecting revenue mobilization. The effect may be positive or negative depending on the means of elimination of trade restrictions. If the tariffs are eliminated, the revenue will fall as the taxable base decreases, but if the tariffs are used to eliminate quotas and exemptions, or if there are reductions in tariff peaks the overall tax collection will increase. Additionally, there is a strong correlation between openness to trade and the size of the government due to the necessity of government-funded social insurance schemes to offset external risks arising from high degrees of dependence of the domestic economy on the rest of the world.

Resource rents as a percent of GDP have been found to have statistically significant relationships with the tax ratio, however different studies as shown in the literature review have found both negative and positive coefficients for various samples and methods of estimation.

Agriculture value added as a percent of GDP has been identified to have a negative relationship with the tax ratio. There are both supply and demand side reasons why such a relationship can hold. From the supply side the agriculture sector, especially if it is primarily subsistence agriculture, is difficult to tax due to the high number of farmers and no taxable income generated. Additionally, from the side of political economy, such a sector is not feasible to tax due to political pressure. From the demand side perspective, the agriculture sector may reduce the provision of public goods which tend to be concentrated in metropolitan areas.

Public goods index and Urbanization are demographic variables. The public goods index is meant to be an alternative for control of government spending on health as it was included in previous studies. Urbanization is included as a control of population spread.

The Public sector corruption index, the Accountability index, and the Rule of law index are meant to be controls of government and institutional quality which was proven to positively affect the tax ratio.

The last variable to be controlled for as described in the dynamic panel models is lagged tax effort. It was first used by Leuthold (1991) and then was used in later papers. There are two basic interpretations, which are Keynesian and Neoclassical. According to Keynesian theory, there exists a loop linking tax collection and public expenditure. A higher level of tax collection will increase public expenditures and thus transform into economic growth which in turn increases tax revenues. This constitutes a direct relationship between tax collection today and in the previous period if taxes are high enough to create said loop. The Neoclassical theory states that a higher level of taxes collected will result in a reduction of economic activity and thus reduce tax revenues. In this way, the theory stipulates that there exists an inverse relationship between taxes today and in the previous year. From a purely statistical point of view, if the coefficient of the dependent variable is approximately equal to one this suggests that the dependent variable is highly conditioned by previously observed value and that explanatory variables have less power to affect the independent variable (Ángeles Castro & Ramírez Camarillo, 2014; Angeles-Castro, 2006).

The literature discussed above provides guidance for the strategy employed in this paper. It is shown that there is a consensus about some economic variables to be included in the estimation of the tax effort, but the set of used variables differs depending on the study and political and demographic variables are still being questioned. Additionally, the method of estimation has evolved from simple regression analysis towards a superior stochastic frontier analysis with the latest development being the true random effects.

3. Estimation strategy

The estimation strategy for this paper is straightforward. The idea is to use the tax effort scores provided by McNabb et al. (2021) for the True Random Effects frontier model which controls for all economic, demographic, and institutional quality as proxied by independent variables. Then, the variables identified in section 2, which have been proven to have statistical significance on the relative revenues collected, are incorporated into regressions for the tax effort scores. In the ideal case, the tax effort score should not be explained by any of those variables, if a variable is found to have statistical significance then it indicates the existence of one of the following two cases: i)

either the frontier is not well estimated and the variable does affect both potential revenue collection and inefficiency parameter, or ii) the frontier is well estimated and the variable only affects the actual tax collection. At this stage, this paper aims to identify some potential excluded variables that may correlate to the tax effort scores and should then be included in further estimation.

To achieve this goal, firstly basic static specifications for the estimation are proposed. All the methods use the same tax effort score as the dependent variable and the same explanatory variables. As argued above, the independent variables included are those which were used by McNabb et al. (2021) to estimate the potential tax, plus those which were described in section 2. The inclusion of the original variables used in the estimation helps to avoid the endogeneity problem as these variables are known to affect the tax effort and therefore will influence the estimations if omitted.

There are three basic potential models which apply to panel data. They are the pooled, fixed effects, and random effects specifications of the ordinary least squares regression model. Therefore, the OLS with fixed effects specification can be represented in the following general vector notation.

$$y_{it} = \alpha_i + \beta X_{it} + \delta Z_{it} + u_{it}$$

Where the y_{it} is the tax effort score of the country i at time t, α_i is the country fixed effect, X_{it} is the set of structural variables with which the tax effort was constructed and Z_{it} is the set of variables identified for further research. β and δ are estimated coefficients.

Fixed effects model controls for country-fixed effects by creating the first difference estimator, thus cancelling unobserved heterogeneity and adjusting the intercept. This specification assumes that some country-specific characteristics are not captured by explanatory variables and are not correlated with the error term. The use of fixed effects is a standard approach to work with panel data as it provides true estimates without the influence of unobserved country-specific heterogeneities which are an important problem when working in cross-country settings. This specification is chosen after contrasting three estimation models based on the tests described below.

First, the F-test is proposed to confirm that the fixed effects specification performs better by rejecting the null hypothesis of additional coefficients equal to zero, and thus the intercept is constant for all the variables $(H_0: \alpha_i = \alpha)$. After, the Breusch Pagan Lagrange Multipliers test is performed to test if random effects specification is appropriate. The test has a null hypothesis that country-specific or time series error variance is zero $(H_0: \sigma_u^2 = 0)$. If both tests reject their respective null hypothesis, meaning that both random and fixed effects perform better than the pooled model, the Hausman test must be performed to select between the two specifications. The null hypothesis of the test is that the regressors and country-specific error term are uncorrelated $(H_0: \rho(X_{it}, \varepsilon_i) = 0)$. If the null hypothesis is rejected fixed effects perform better than random effects specification.

Finally, robust standard errors are estimated to adjust for heteroskedasticity and serial autocorrelation in error terms. If ordinary least squares assumptions about the distribution of the error term, i.e. white noise, are not satisfied and errors are not homoscedastic or serially uncorrelated, then inferences are misleading. Standard errors are likely to be underestimated and biased, thus regressions would not provide a correct assessment of the statistical significance of estimated coefficients of explanatory variables. The use of robust standard errors as described by Rogers (1994) will give meaningful results under the assumption that countries are independent

from one another but may be dependent on previously observed values of themselves. This allows me to produce consistent standard errors and alleviate heteroskedasticity and serial autocorrelation problems in panel data.

In summary, the tax effort scores will be taken as independent variables, and ordinary least squares estimation with fixed effects and robust standard errors will be conducted. The estimation from this model will be reported in section 5. Additionally, this estimation will be carried out for four sub-sets of observations. The whole sample will be split into four sub-sets based on income, the groups are "Low Income", "Lower-Middle Income", "Upper-Middle Income" and "High Income".[1] These estimates are done to check for heterogeneities based on the achieved development of countries in the frontier estimations, and this categorization is supported by the literature as income is a proxy for the level of development. In turn, the stage of development was proven to have effects on statistical significance and signs of several variables as discussed in section 2.

Next, by using the evidence from these estimations I will replicate, to the best of my ability, the construction of the stochastic frontier as described in McNabb et al. (2021) and then verify the definite effect of proposed variables on the potential tax collections and discuss the implications of my estimations on the reported tax effort scores.

Although this strategy is simple it allows for the identification and discussion of potential issues with frontier estimation and the proposal of new variables to inclusion into such estimations to report consistent and statistically more robust scores. In other words, variables identified in this paper are potential sources of endogeneity for tax effort estimations and must be tested against the frontier to be either included or omitted based on their effect on the tax capacity.

4. Data

The data used in this work is extracted from several sources. In particular, the data on taxes excluding social contributions, tax effort scores, the dummy variable identifying countries reporting revenue statistics at the general government level, total resource revenue, and grants are obtained from the UNU-WIDER Government Revenue Dataset (see McNabb et al., 2021; UNU-WIDER, 2023) which provides comprehensive data for a sample of 196 countries at all development and income levels for the period since 1980 until 2022. The data on the shadow economy index comes from Medina & Schneider (2018), which provides an index of shadow economies for 158 countries and a period from 1991 to 2015. The data on economic variables such as central government debt, exports, imports, trade, foreign direct investment, agriculture share of GDP, GDP per capita adjusted for purchasing power parity, inflation, and share of urban population all come from the world development indicators database (World Bank, 2023). Additionally, the data on the region and income of the country for the period is retrieved from the same World Bank dataset. Next, data on the particularistic or public goods, accountability index, public corruption indexes, and the rule of law come from the varieties of democracy (V-Dem) dataset (Michael Coppedge et al., 2024). Finally, gross fixed capital formation comes from the OECD database (OECD, 2024), thus it has a very limited number of observations. Moreover, there is a poor coverage of investment rates outside developed countries and therefore this variable as well as central government debt are not going to be used in the main estimations. However, estimations of section 5 for the entire sample are provided in Appendix I for completeness.

[1] The four categories are used as defined by The World Bank. The main criteria to allocate a county into an income group is based on a threshold for gross national income per capita, in U.S. dollars. Thresholds were established in 1989 and are adjusted with inflation (World Bank, n.d.-a).

As this paper is based on the tax effort scores and revenue statistics provided in the UNU-WIDER Government Revenue Dataset, I employ almost the same sample of countries in order to reach meaningful conclusions.[2] In this way the full sample for which the data is available for all the variables is composed by 1936 observations including 108 countries across the period from 1991 to 2015. The summary statistics are presented in *Table 1*.

Table 1: Summary statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Taxes excluding social security contributions	1936	17.542	8.47	.607	49.369
Total Resource Revenue	1936	2.52	7.292	-.029	65.268
Tax effort (TREz)	1936	.833	.078	.217	.974
Dummy if General Govt data	1936	.514	.5	0	1
Grants	1936	1.111	2.716	0	35.634
Exports of goods and services	1936	41.128	28.327	4.902	228.994
Imports of goods and services	1936	44.335	24.224	6.97	208.333
Agriculture share in GDP	1936	11.791	11.563	.033	61.416
GDP per capita, PPP	1936	15574.29	16702.178	498.338	107859.69
Urban population	1936	56.565	22.818	5.491	100
Inflation, consumer prices	1936	8.238	29.951	-11.686	1058.374
Foreign direct investment	1936	4.369	7.362	-57.532	86.479
Central government debt	778	57.837	35.129	-1.171	194.584
Particularistic or public goods	1936	.892	1.105	-2.648	3.068
Accountability index	1936	.881	.815	-1.263	2.069
Public sector corruption index	1936	.418	.325	.001	.968
Rule of law index	1936	.63	.313	.032	.999
Shadow economy index	1936	28.48	12.944	6.16	71.27
Gross fixed capital formation	744	3.575	8.95	-38.903	51.078

Source: own elaboration.

5. Empirical analysis

I) Graphical analysis

The first step in the analysis is to plot the tax effort score against the variables under consideration and obtain preliminary results. Additionally, graphs are presented for subsamples conditioned on the income group the country falls to in that year. These graphs are presented in *Figures 2-13*.

The first finding is a positive relationship between foreign direct investment and the tax effort score. Although this relationship is not straightforward since countries are concentrated on the FDI axis around zero, there are several observations suggesting a positive relationship. This is demonstrated through observations that rely more heavily on FDI inflow and simultaneously achieve larger values of tax effort, thus moving along the fitted values line. However, this relationship is heterogeneous among groups of countries by income. Low-income countries observe much larger increases in their tax effort with FDI increases, while this relationship appears insignificant for countries in other higher-income groups.

[2] Differences arise from the availability of other data employed from other sources.

Second, the size of the shadow economy appears to have a negative relationship with the tax effort. This fact is supported by the previous literature as well as the general economic relationship between the shadow economy, which doesn't contribute to the collection of tax revenues and therefore will reduce the tax effort. Unexpectedly, if tax effort is plotted by income group, the low-income countries show a positive relationship between the two variables, while countries in other income groups display a homogeneous negative relationship. The fact that low-income countries experience a positive relationship is in opposition to the current literature and generally is not possible. Although the graphical analysis suggests this positive relationship in the next section, when applying the appropriate regression, the coefficient of the shadow economy is found to be negative across all income levels and therefore this apparent discrepancy may arise due to two possible sources. The first one is the fact that there may be unobserved country characteristics that are not shown in the graph and that do influence the tax effort and create this positive relationship. The second one is a possibility that there is another variable, omitted from the graph which is positively correlated with both tax effort and shadow economy, thus creating this disparity relationship while the true effect is negative but obscured. It is also important to mention that possible discrepancies may arise since the shadow economy and many other variables are proxied by an index that may be based on unreliable economic data. Concerns about data quality are most severe for the countries with the lowest incomes lacking tools and statistical institutions.

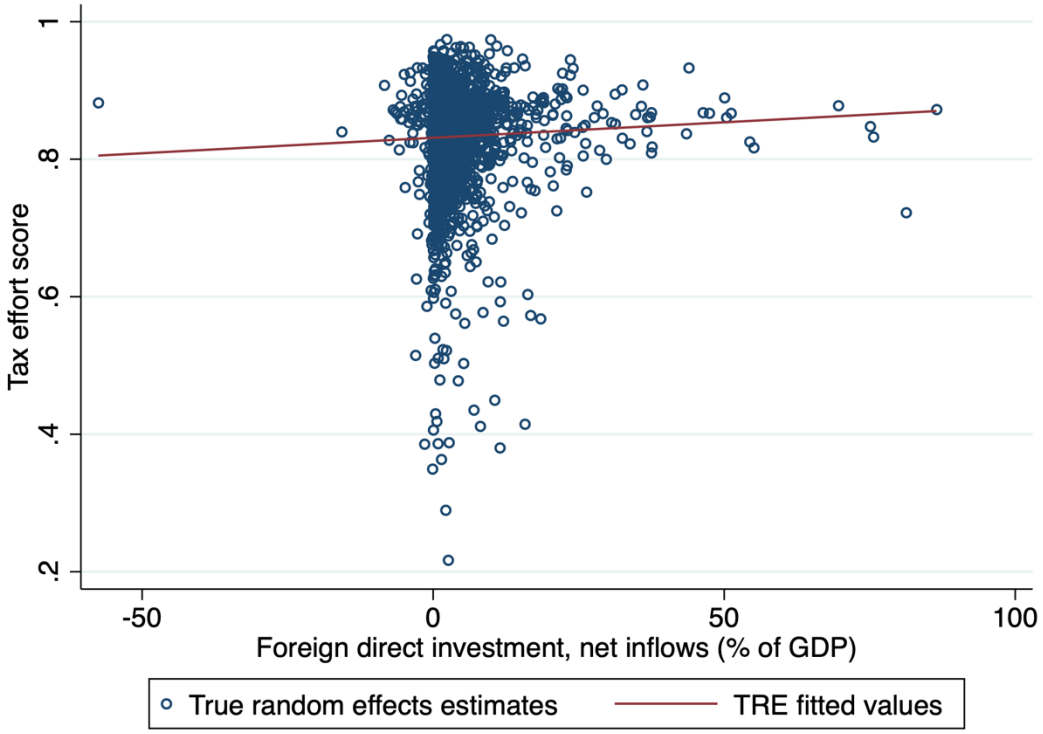
Next, grants do not display any relationship with tax effort. Nevertheless, if disaggregated by income group higher values of grants appear to have a positive relationship with revenues and no significance for both middle-income groups. High-income countries observe a negative relationship which is consistent with the literature, but this conclusion cannot be inferred from the graph alone as those countries are not typical receivers of grants. From the logic outlined before, it follows that countries with lower incomes are relying on grants for productive investments which contribute to increases in tax bases, while higher-income countries use grants to finance their current consumption.

Central government debt covers a low number of observations, especially for low-income countries, with a total of 778 observations and no apparent relationship with tax effort. In the sub-sample figures, the relationship is positive for low-middle-income countries and null for other income groups.

Investment as measured with the gross fixed capital formation also has a limited number of observations with a total of 744, and no data for low-income countries. In the general dataset as well for upper-middle- and high-income countries, which comprise most of the observations, the relationship appears to be positive. This is the expected relationship in the literature.

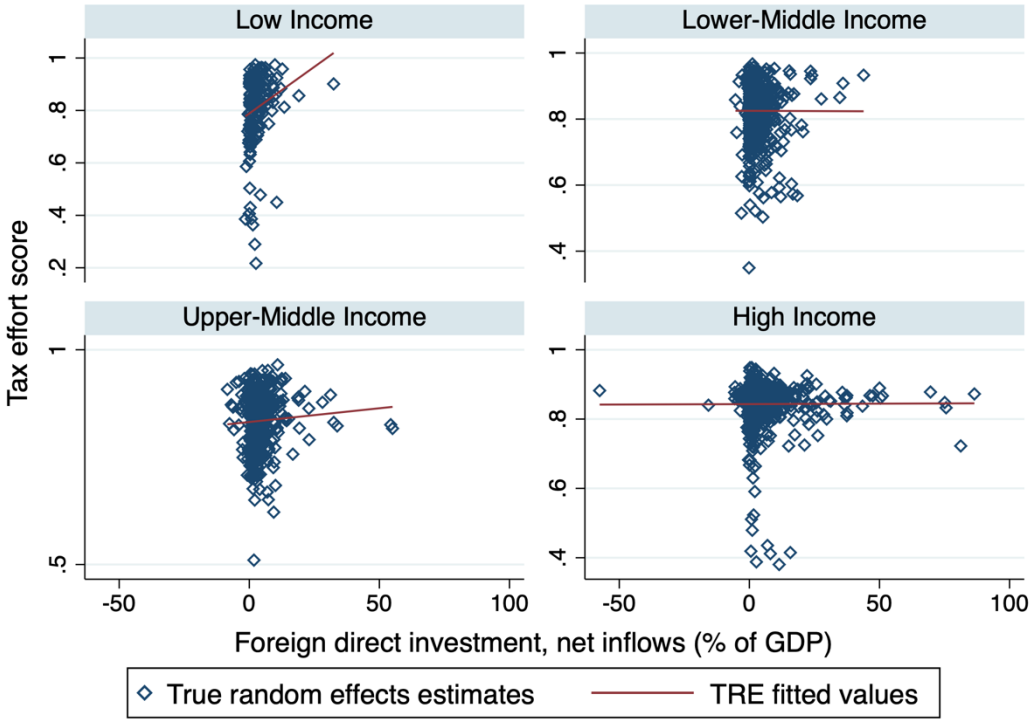
Finally, the inflation rate has a negative relationship with the tax effort scores. Although most of the observations are concentrated around 8% there are many countries with extreme inflation rates. In the sub-sample figure, low-income and lower-middle-income countries display a strong negative relationship while the upper-middle- and high-income countries' tax effort scores do not display the same evolution and suggest no relationship with inflation.

Figure 2: Tax effort over the FDI.



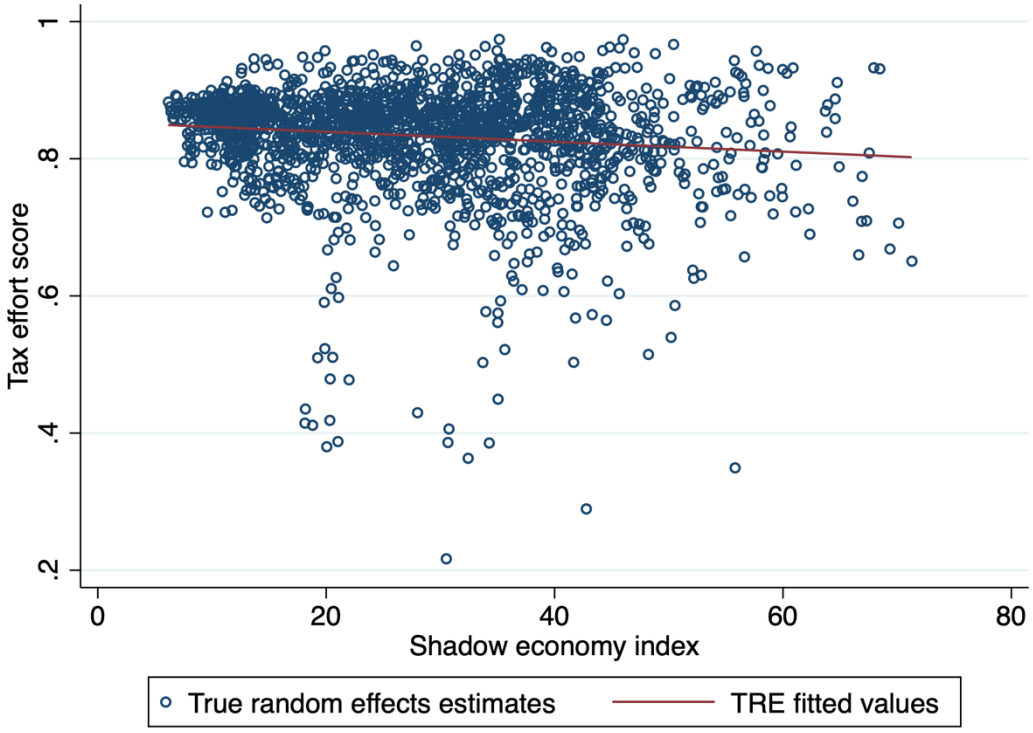
Source: own elaboration.

Figure 3: Tax effort over the FDI by income group.



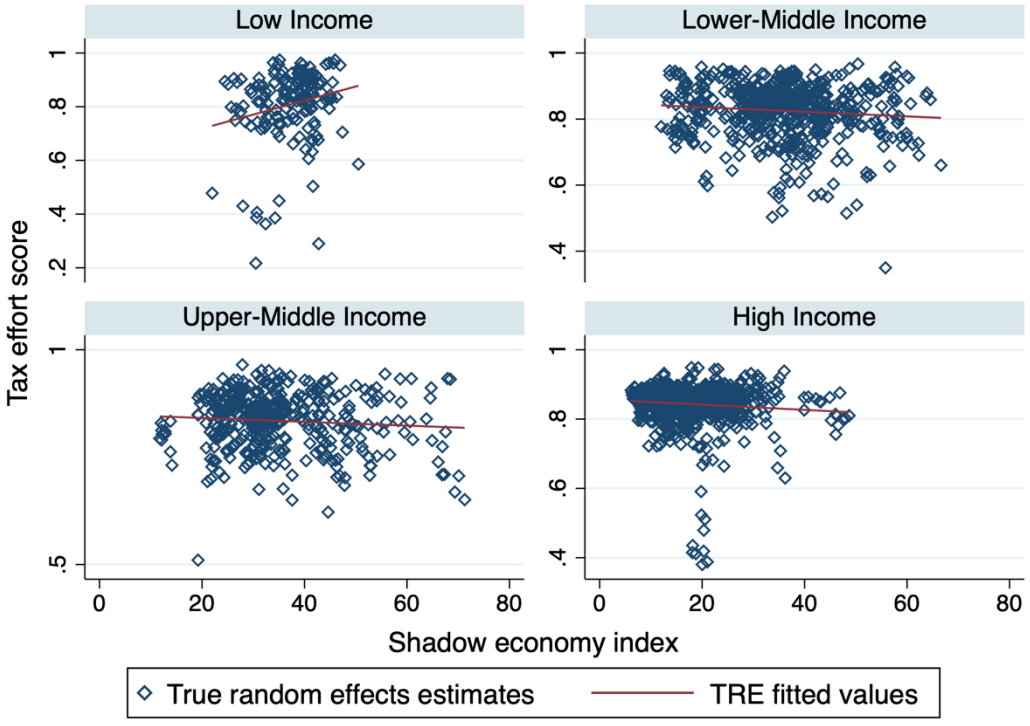
Source: own elaboration.

Figure 4: Tax effort over the shadow economy index.



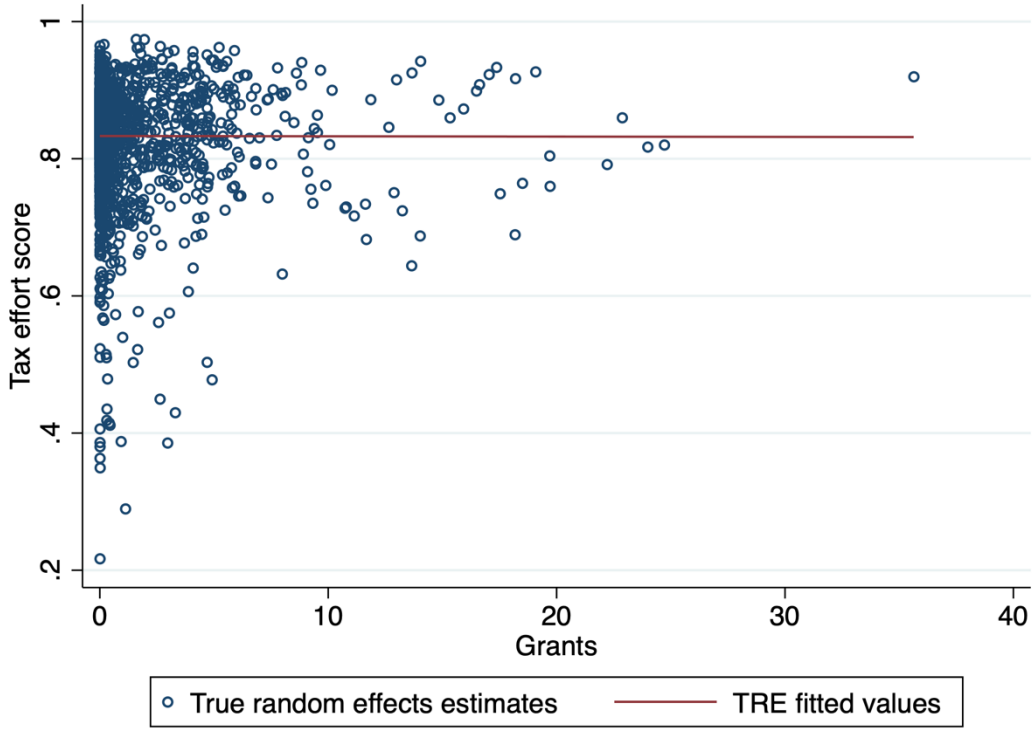
Source: own elaboration.

Figure 5: Tax effort over the shadow economy index by income group.



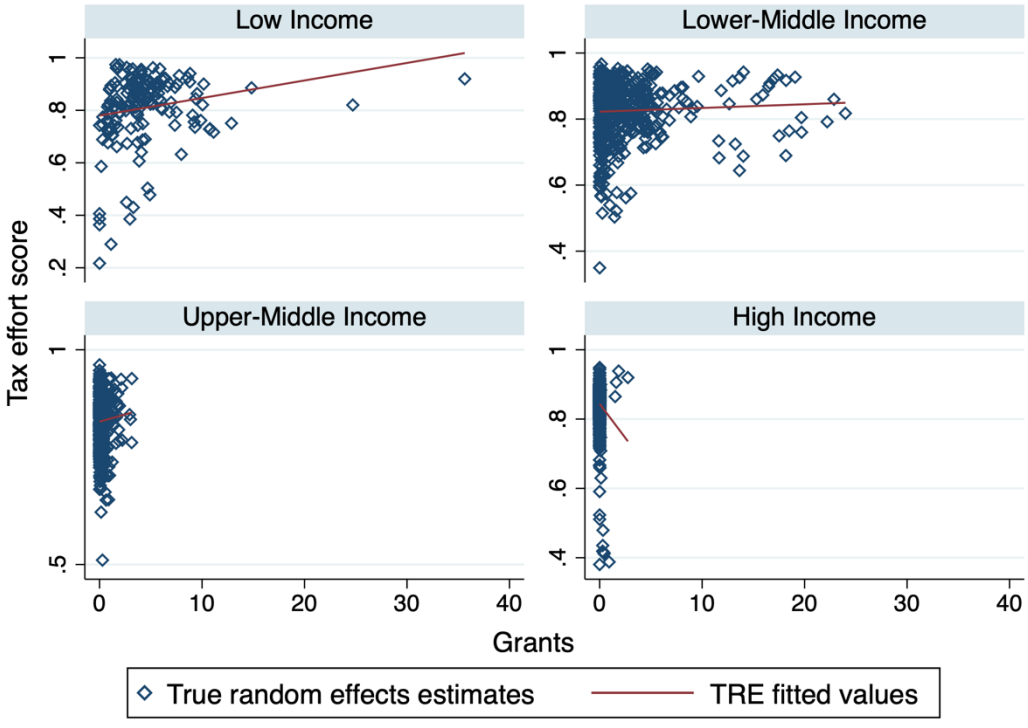
Source: own elaboration.

Figure 6: Tax effort over the grants.



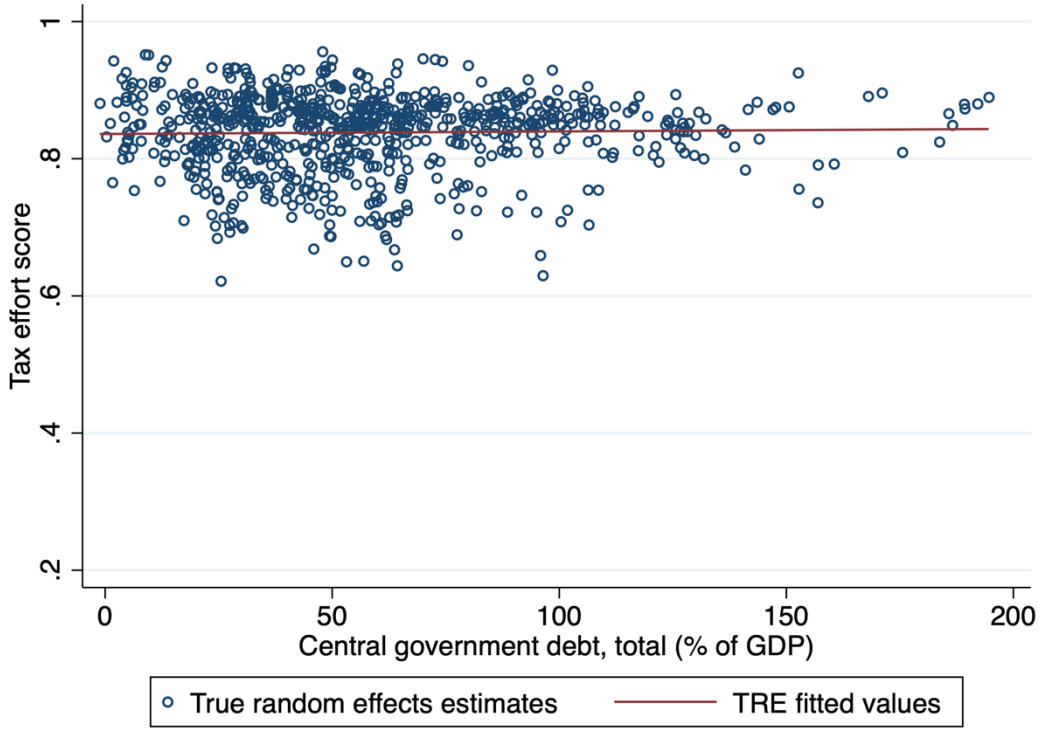
Source: own elaboration.

Figure 7: Tax effort over the grants by income group.



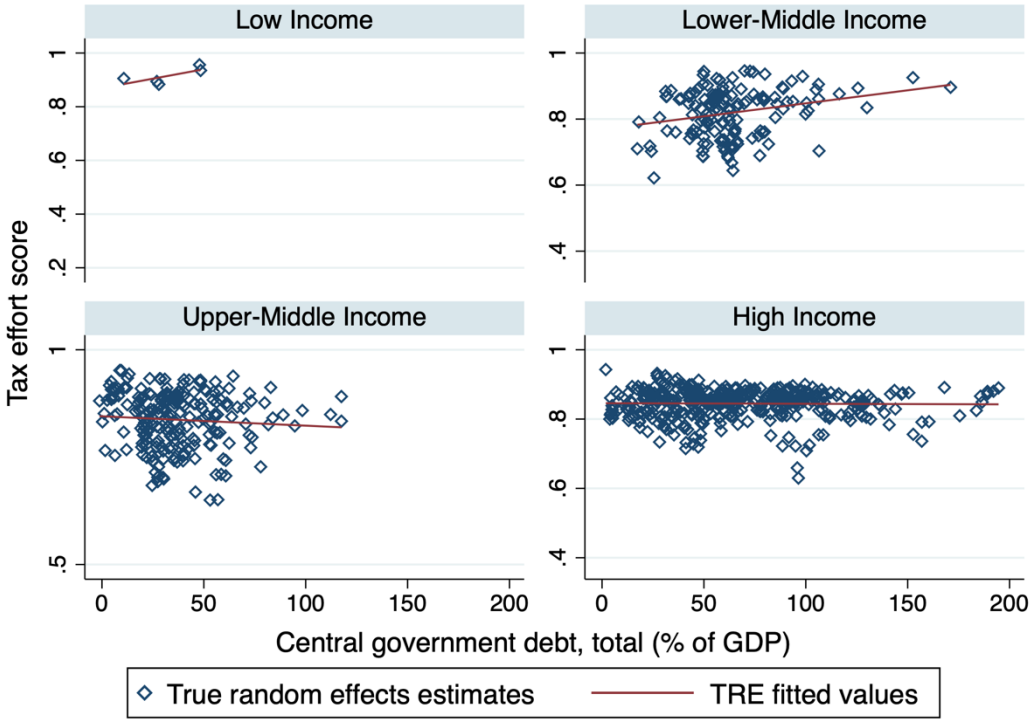
Source: own elaboration.

Figure 8: Tax effort over debt.



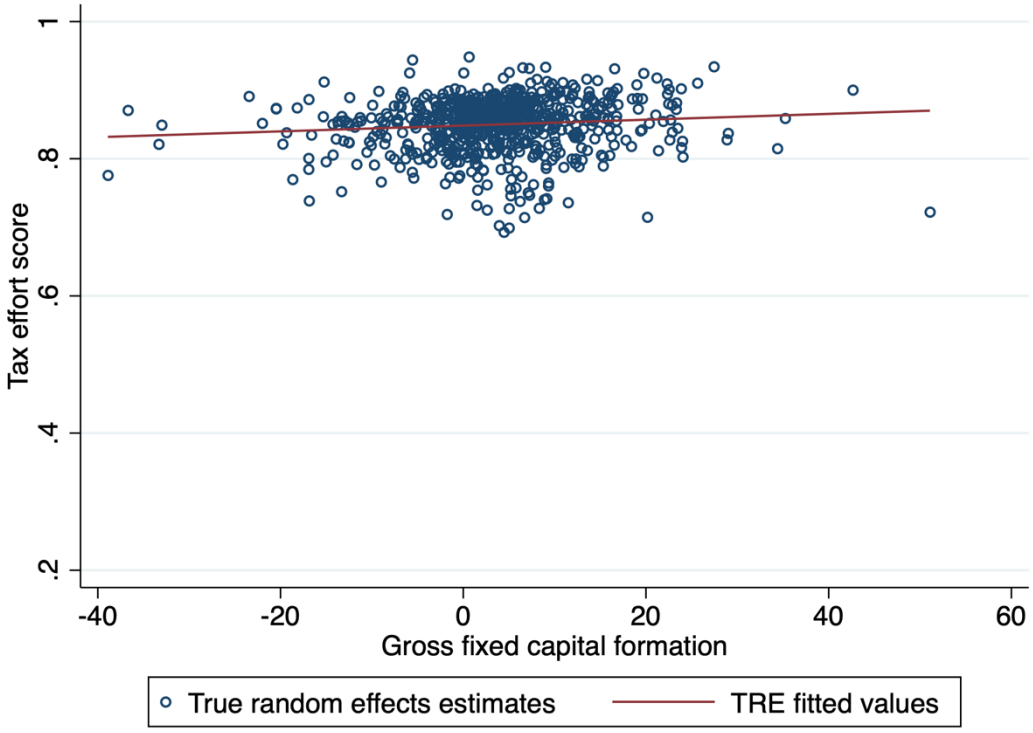
Source: own elaboration.

Figure 9: Tax effort over debt by income group.



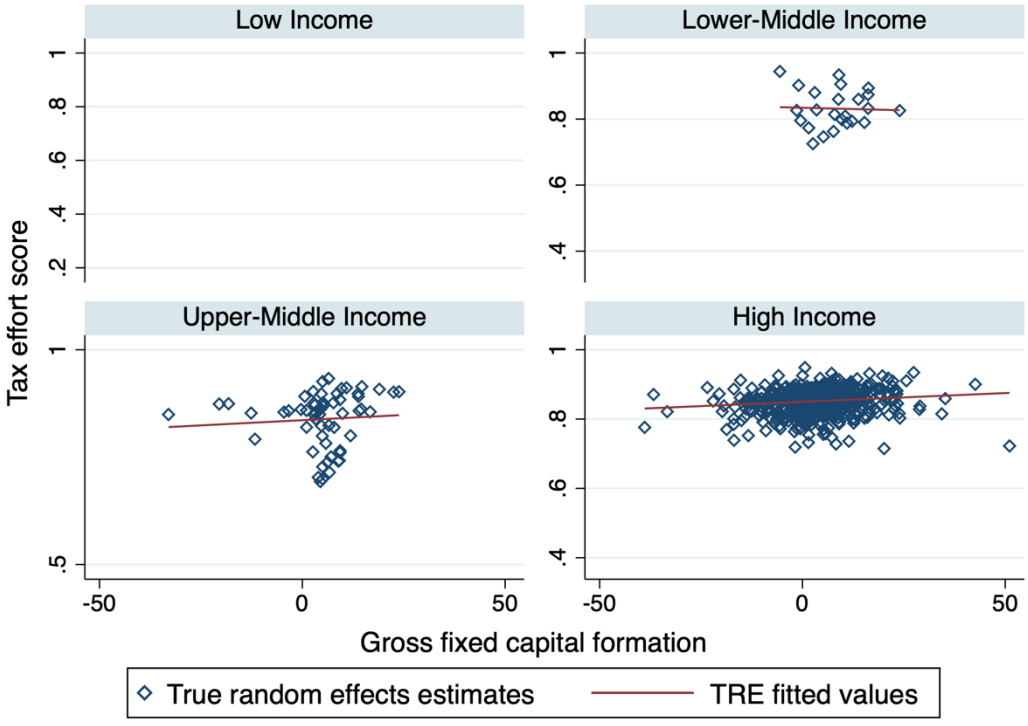
Source: own elaboration.

Figure 10: Tax effort over investment.



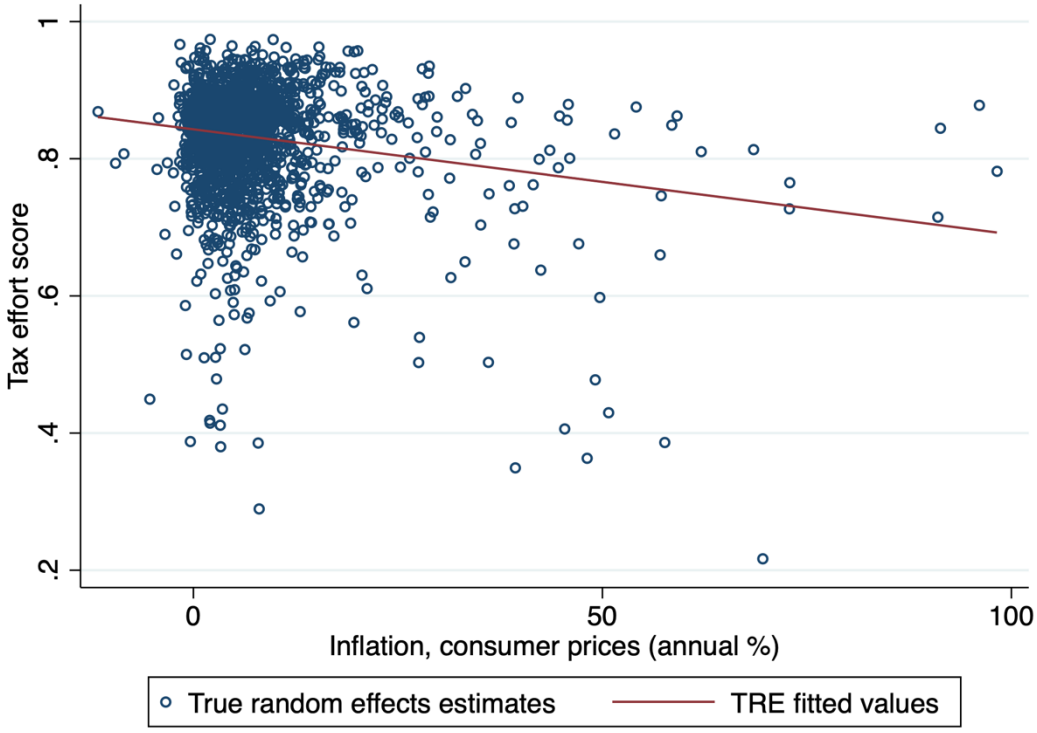
Source: own elaboration.

Figure 11: Tax effort over investment by income group.



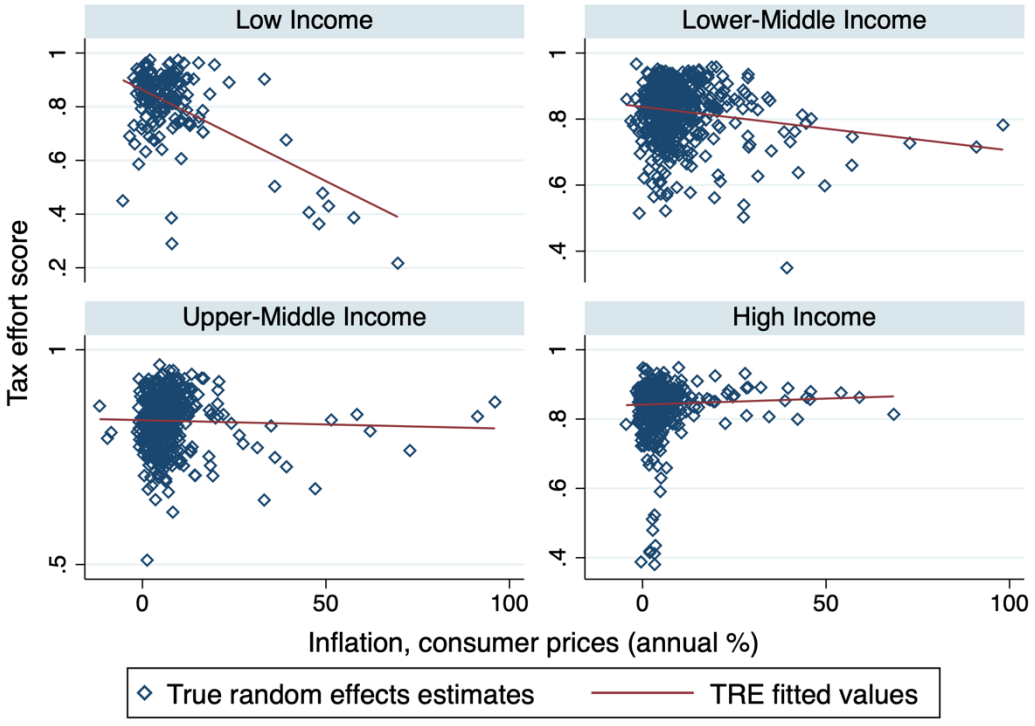
Source: own elaboration.

Figure 12: Tax effort over inflation rate.



Source: own elaboration.

Figure 13: Tax effort over inflation rate by income group.



Source: own elaboration.

II) Baseline regression

The results from this estimation are meant to determine if any of the variables identified above are capable of distorting estimated tax effort scores. Since both random effects and fixed effects models reject their corresponding null hypothesis the Hausman test is performed. The test rejects the null of unsystematic differences in coefficients and thus, fixed effects are more appropriate. The results of the Hausman test are presented in *Table 2*.

Table 2: Hausman test.

Null Hypothesis	Difference in coefficients not systematic
chi2(12)	47.47
Prob > chi2	0.0000

Source: own elaboration.

As argued in section 3, the best fitting regression to apply for the tax effort as the dependent variable is the Country Fixed Effects with Robust Standard Errors. The results of this baseline regression are presented in *Table 3*.

Table 3: Fixed effects regressions with robust standard errors for the tax effort.

	(1) Tax effort (TREz)
GDP per capita, PPP (current international $)	-0.000000897
	(-0.66)
GDP per capita, PPP (current international $) squared	8.65e-12
	(0.70)
Trade (%GDP)	-0.000206
	(-0.73)
Total Resource Revenue	0.000282
	(0.11)
Agriculture, forestry, and fishing, value added (% of GDP)	-0.00211
	(-0.90)
Particularistic or public goods	-0.00789
	(-0.44)
Urban population (% of total population)	0.000176
	(0.11)
Public sector corruption index	-0.168
	(-1.13)
Accountability index	0.0929***
	(2.74)
Rule of law index	-0.297*
	(-1.69)
Shadow economy index	-0.000218
	(-0.12)
Grants	0.00637**
	(2.34)
Foreign direct investment, net inflows (% of GDP)	0.000262
	(0.79)

Inflation, consumer prices (annual %)	-0.0000110
	(-0.16)
Constant	1.055***
	(4.79)
Observations	1936
Adj. R-squared	0.0556

Note: *t* statistics in parentheses and based on robust S.E.; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; *Source: own elaboration.*

The results presented in this table show that, in line with the expectations, all baseline variables except for accountability and rule of law indexes are not significant. The accountability index and rule of law index which are found to be statistically significant are omitted from the frontier estimation available since authors employ three different specifications to control for institutional quality but report only one. Nevertheless, this result is expected as those variables are shown to be statistically significant for the estimations and thus will be correlated with the reported scores. Despite appearing relevant in the graphical analysis, coefficients of all the additional variables except for grants are not found to be statistically significant. Thus, grants received are shown to be potentially distorting for reported tax effort scores after being neglected by McNabb et al. (2021). Such finding, preliminary demonstrates that reported tax effort scores will not quantitatively change if all the additional variables except grants are introduced in the frontier design. Nevertheless, close attention must be paid if the sample is split into income groups. Sub-sample regressions are presented in *Table 4*.

Table 4: Fixed effects regressions with robust standard errors for tax effort by Income.

	Low income	Lower-Middle Income	Upper-Middle Income	High Income
GDP per capita, PPP	-0.000257	-0.0000194	-0.00000859	0.000000646
	(-0.62)	(-1.26)	(-1.16)	(0.57)
GDP per capita, PPP squared	3.97e-08	7.09e-10	1.13e-10	3.04e-13
	(0.34)	(0.90)	(0.65)	(0.03)
Trade (%GDP)	0.000286	-0.000281	-0.0000337	-0.000906***
	(0.45)	(-0.43)	(-0.07)	(-3.78)
Total Resource Revenue	0.0443***	-0.000711	-0.000359	-0.00136
	(21.44)	(-0.13)	(-0.15)	(-0.61)
Agriculture, value added (% of GDP)	-0.000575	-0.00176	-0.00133	0.00302
	(-0.13)	(-0.46)	(-0.47)	(0.90)
Particularistic or public goods	0.0416	-0.0368	0.00167	-0.0276***
	(0.92)	(-1.65)	(0.08)	(-2.74)
Urban population (% of total population)	0.00199	0.00393	-0.00203	0.0000706
	(0.35)	(1.42)	(-1.04)	(0.04)
Public sector corruption index	0.163	-0.159	0.369*	-0.324*
	(0.88)	(-1.13)	(1.83)	(-1.91)
Accountability index	0.156***	0.137***	-0.128**	0.0563
	(3.59)	(4.17)	(-2.44)	(1.04)
Rule of law index	-0.0383	-0.357*	0.644**	-0.328*
	(-0.24)	(-2.02)	(2.25)	(-1.75)
Shadow economy index	-0.00406	-0.000674	-0.00794***	-0.00356***
	(-0.94)	(-0.25)	(-2.89)	(-2.98)
Grants	0.00320*	0.00499**	0.0174	0.00501

	(1.96)	(2.06)	(1.28)	(0.39)
Foreign direct investment (%GDP)	0.00372	0.000658	0.000718	-0.00000236
	(1.23)	(0.44)	(0.71)	(-0.01)
Inflation, consumer prices (annual %)	-0.00539***	-0.0000843	-0.0000176	0.000197
	(-7.18)	(-0.28)	(-0.52)	(1.06)
Constant	0.998*	1.015***	0.853***	1.258***
	(2.14)	(3.40)	(3.21)	(5.54)
Observations	160	585	416	775
Adj. R-squared	0.590	0.114	0.118	0.135

Note: *t* statistics in parentheses and based on robust S.E.; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; *Source: own elaboration.*

However, even in this simple design, the reported tax effort presents significant issues in sub-sample regressions. Out of the basic variables total resource revenue is found to be statistically significant for low-income countries. Then, trade and public goods indexes are found to be significant for high-income countries. Additionally, the public sector corruption index is significant for upper-middle-income countries and high-income countries. The accountability index is statistically significant for all income levels except for high-income. The rule of law index is statistically significant for all income groups except low-income countries. Corruption, accountability, and rule of law indexes aim to capture the same effect of institutional quality, and therefore sign reversals and changes in significance may appear depending on their interaction. Signs of these three variables tend to change depending on the combination used for estimations (see McNabb et al., 2021). The fact that the underlying variables present such high degrees of statistical significance in the sub-sample regressions poses a question about the reliability of general estimations. From the augmented variables, the shadow economy index is found to be statistically significant for upper-middle- and high-income countries. This finding is supported by the previous literature and confirms the existence of the possible omission of this variable in estimations. Grants as shown from the general sample are found to be statistically significant in the low-income sample at 10% and lower-middle-income countries at 5% both with consistent positive coefficients. Nevertheless, grants are not found to be significant for other income groups, suggesting once again the importance of sub-sample analysis. This variable is known to have idiosyncratic effects on each country and is usually targeted at developing nations, thus corroborating the conclusions drawn from the graphical analysis. Foreign direct investment does not exhibit any significant relationship for any of the income groups. Finally, the inflation rate has a significant coefficient for low-income countries, which is confirmed in the graphical analysis as well as discussed in the literature. Therefore, all the variables discussed in the literature are found to be explanatory variables for certain income groups. These findings suggest that reported tax effort scores and potential tax revenue estimates may be inaccurate. The question of whether the variables presented above are relevant for frontier estimation is answered in section 6.

In conclusion, this simple test using regression analysis presents an argument in favour of the tax effort scores provided by McNabb et al. (2021) since they are not explained and thus are not statistically correlated with any of the new potential explanatory variables in the general sample. In this way, they can be applied in research focused on international comparisons. However, there is an important underlying assumption. All countries are subject to the same production frontier, meaning that regardless of its level of development a country can achieve the same level of potential tax collection as any other.

Nevertheless, there are concerns about the use as well as the accuracy of the reported estimates if samples are different or focused on a subset of countries. Thus, the evidence from the regression above presents a case against the use of the tax effort scores in samples highly biased to a particular income group, especially those focused on low-income countries. This is a potential issue with

reported tax effort scores as it was proven in the previous literature that the level of development has important consequences for the coefficients of the estimations. Although, this fact is usually ignored in the estimations of tax effort scores. Because of the high statistical significance and different variables relevant to the tax effort score at each income level, it is important to consider peer group countries by their level of development or income to produce relevant estimates for the sample and income group considered.

6. Stochastic frontier analysis

The next step in this paper is to use the variables discussed above to replicate the process followed by McNabb et al. (2021), create the stochastic frontier with true random effects, and report coefficients as well as final tax effort scores (see also Kumbhakar & Lovell, 2000). Although the tests above were conducted to find the variables that may affect the frontier estimation, statistically insignificant variables such as most of the underlying are known to be important for the stochastic frontier construction. Even though in the general sample shadow economy, FDI, and inflation are not found to be significant, they are included in the frontier estimation as control variables. The results of the frontier are presented below in *Table 5*.

Table 5: Stochastic Frontier Analysis results.

	(1) Log taxes excluding social contributions	(2) Log taxes excluding social contributions
=1 if General Govt data	0.309***	0.316***
	(16.77)	(16.95)
GDP per capita, PPP	1.432***	1.504***
	(14.20)	(13.63)
GDP per capita, PPP squared	-0.0783***	-0.0825***
	(-13.79)	(-13.46)
Trade (%GDP)	0.120***	0.137***
	(7.36)	(7.91)
Total Resource Revenue	0.0752***	0.0774***
	(8.03)	(8.29)
Agriculture, value added (% of GDP)	-0.0935***	-0.0918***
	(-4.83)	(-4.75)
Particularistic or public goods	-0.0124	-0.0233
	(-0.72)	(-1.32)
Urban population (% of total population)	-0.108***	-0.102***
	(-4.05)	(-3.80)
Public sector corruption index	-0.607***	-0.619***
	(-5.74)	(-5.74)
Accountability index	-0.0430	-0.0422
	(-1.42)	(-1.37)
Rule of law index	0.313**	0.278*
	(2.15)	(1.83)
Shadow economy index		-0.0467*
		(-1.91)
Grants		0.0242
		(1.50)

Foreign direct investment (% of GDP)		-0.0214***
		(-3.02)
Inflation, consumer prices (annual %)		0.00133
		(0.18)
Constant	-3.386***	-3.602***
	(-8.61)	(-8.22)
lnsig2v		
Constant	-3.272***	-3.312***
	(-51.37)	(-50.21)
lnsig2u		
Constant	-2.641***	-2.619***
	(-34.34)	(-34.10)
N	1936	1936

Note: *t* statistics in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; *Source: own elaboration.*

Firstly, column 1 presents the results of the estimation of the baseline frontier that replicates the estimates presented by McNabb et al. (2021) for the available sample of countries. In general, coefficients in the baseline frontier reported are close but not identical to those obtained by McNabb et al. (2021). Nevertheless, coefficients of total resource revenue, public goods index, urbanization, and accountability index were expected to have opposite signs from those obtained. Additionally, the public goods index is not reported as statistically significant while it was found to be relevant. The differences in coefficients and signs reversals are explained by the fact that the sample used to conduct all these estimations is significantly reduced to about 50% of the one for which tax effort scores are reported and to 31% of the total sample for which total taxes are available. Therefore, this reinforces the point that these estimations are not independent of the sample and must be compared only to countries in a similar state of development.

The inclusion of relevant augmented variables does not significantly influence reported coefficients and their significance levels (column 2). Out of the basic variables, only the rule of law becomes less significant if new variables are included. Therefore, this extended definition of the frontier refines final tax effort estimates and makes them more realistic. This is the case because the new model includes statistically significant variables that do not affect the prediction power and magnitude of the ones previously used, thus providing better estimates of inefficiency. As mentioned before, grants are not found to be statistically significant to explain the log of total taxes excluding social contributions and thus will not affect the frontier estimation, although the variable has a statistical correlation with tax effort score. Shadow economy, as discussed in the previous section, is found to have a statistically negative effect on tax collection in the general sample although it does not have explanatory power for the tax effort. In the same way, foreign direct investment, which was not flagged as relevant in the general regression nor sub-sample regression is shown to be statistically significant at a 1% significance level and reduces the total taxes collected. This was the expected effect from the literature review, as the high level of FDI is correlated with lower relative tax rates in the receiving country. Finally, inflation is not found to be important in the general sample frontier.

Therefore, by using both estimations presented in Table 5, as well as original estimates by McNabb et al. (2021), the following tax effort scores are reported.[3]

[3] The tax effort score is reported once for each country in the most recent available year in the dataset.

Table 6: Estimated tax effort scores.

Country	Year	Tax effort McNabb et al. (2021)	Tax effort baseline	Tax effort augmented
Albania	1999	0.838	0.833	0.820
Angola	2005	0.838	0.952	0.949
Austria	1997	0.861	0.875	0.866
Azerbaijan	2001	0.835	0.869	0.880
Bahrain	2013	0.380	0.109	0.107
Bangladesh	2004	0.877	0.726	0.733
Belarus	2015	0.774	0.890	0.889
Belgium	2005	0.864	0.875	0.880
Bhutan	2010	0.923	0.730	0.696
Bosnia and Herzegovina	2007	0.921	0.916	0.920
Botswana	2014	0.858	0.877	0.877
Bulgaria	1997	0.815	0.890	0.893
Burkina Faso	2000	0.793	0.853	0.853
Cabo Verde	2004	0.862	0.887	0.886
Cameroon	2004	0.819	0.850	0.840
Canada	2008	0.859	0.881	0.882
Chad	2005	0.386	0.453	0.415
Chile	2010	0.844	0.745	0.747
China	2009	0.804	0.873	0.868
Comoros	2013	0.860	0.716	0.678
Congo, Dem. Rep.	2015	0.955	0.908	0.909
Congo, Rep.	2000	0.515	0.282	0.265
Costa Rica	2008	0.910	0.798	0.808
Cote d'Ivoire	2008	0.876	0.803	0.801
Cyprus	1993	0.708	0.855	0.859
Czech Republic	1996	0.852	0.779	0.769
Denmark	2013	0.868	0.944	0.944
Dominican Republic	1996	0.831	0.804	0.793
Egypt, Arab Rep.	1992	0.932	0.902	0.896
El Salvador	2012	0.902	0.906	0.910
Equatorial Guinea	2010	0.757	0.467	0.470
Estonia	2006	0.839	0.702	0.708
Eswatini	2010	0.608	0.883	0.888
Ethiopia	2014	0.894	0.859	0.859
Fiji	2007	0.838	0.903	0.910
Finland	1991	0.887	0.913	0.910
France	2006	0.865	0.883	0.884
Gabon	1991	0.869	0.896	0.894
Georgia	2003	0.789	0.858	0.868
Germany	2000	0.852	0.780	0.789
Ghana	2011	0.887	0.794	0.805
Greece	1998	0.828	0.863	0.863
Guinea	2005	0.828	0.851	0.854
Guinea-Bissau	2000	0.933	0.850	0.830

Haiti	2011	0.904	0.783	0.782
Honduras	2009	0.801	0.827	0.820
Hungary	2008	0.867	0.843	0.855
Iceland	2012	0.756	0.917	0.919
India	2009	0.763	0.835	0.822
Indonesia	2004	0.914	0.872	0.870
Iran, Islamic Rep.	2000	0.795	0.345	0.332
Ireland	2007	0.925	0.850	0.856
Israel	2014	0.868	0.879	0.882
Italy	2010	0.875	0.912	0.915
Jamaica	1996	0.801	0.864	0.865
Japan	2005	0.840	0.713	0.696
Jordan	2013	0.742	0.846	0.831
Kazakhstan	1999	0.622	0.685	0.694
Kenya	2009	0.843	0.905	0.906
Kuwait	2013	0.511	0.091	0.090
Lao PDR	1995	0.562	0.331	0.331
Latvia	2012	0.829	0.728	0.722
Lithuania	2010	0.732	0.616	0.614
Luxembourg	2010	0.878	0.869	0.880
Madagascar	2011	0.844	0.899	0.903
Malaysia	2014	0.802	0.824	0.821
Mali	2013	0.818	0.908	0.907
Mauritania	1994	0.876	0.853	0.845
Mauritius	2007	0.771	0.753	0.754
Mexico	2014	0.871	0.774	0.779
Moldova	2005	0.847	0.909	0.911
Mongolia	2015	0.769	0.817	0.791
Myanmar	2014	0.889	0.624	0.637
Namibia	2004	0.757	0.909	0.912
Nepal	1999	0.708	0.748	0.746
Netherlands	2008	0.866	0.820	0.826
New Zealand	1997	0.858	0.921	0.921
Niger	2015	0.935	0.905	0.911
Nigeria	2003	0.853	0.548	0.553
Norway	1995	0.807	0.859	0.857
Pakistan	1996	0.892	0.855	0.855
Papua New Guinea	2002	0.847	0.872	0.862
Paraguay	1995	0.943	0.844	0.832
Poland	2014	0.804	0.756	0.757
Portugal	1998	0.857	0.795	0.804
Romania	2004	0.821	0.875	0.884
Rwanda	2003	0.891	0.898	0.897
Sierra Leone	2008	0.811	0.890	0.893
Singapore	1998	0.816	0.596	0.588
Slovak Republic	2003	0.822	0.676	0.660
Slovenia	2011	0.856	0.757	0.751
Spain	1997	0.822	0.782	0.788
Suriname	2007	0.908	0.842	0.805

Sweden	2008	0.868	0.915	0.916
Switzerland	2007	0.876	0.751	0.741
Tajikistan	2008	0.834	0.895	0.897
Tanzania	2015	0.820	0.820	0.824
Thailand	2011	0.851	0.798	0.798
Togo	1993	0.586	0.775	0.780
Tunisia	1997	0.864	0.866	0.865
Turkey	2013	0.891	0.874	0.872
Uganda	1996	0.920	0.844	0.851
United Arab Emirates	2013	0.943	0.725	0.721
United Kingdom	2009	0.850	0.848	0.838
Uruguay	2008	0.846	0.725	0.744
Vietnam	1999	0.820	0.634	0.618
Zambia	2007	0.804	0.824	0.834
Zimbabwe	2012	0.870	0.936	0.940

Source: own elaboration.

Finally, I produce sample statistics as well as histograms to present the total outlook on the estimations.

Table 7: Tax effort descriptive statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Tax effort McNabb et al.	1936	.833	.078	.217	.974
Tax effort baseline	1936	.796	.141	.091	.964
Tax effort augmented	1936	.794	.143	.09	.964

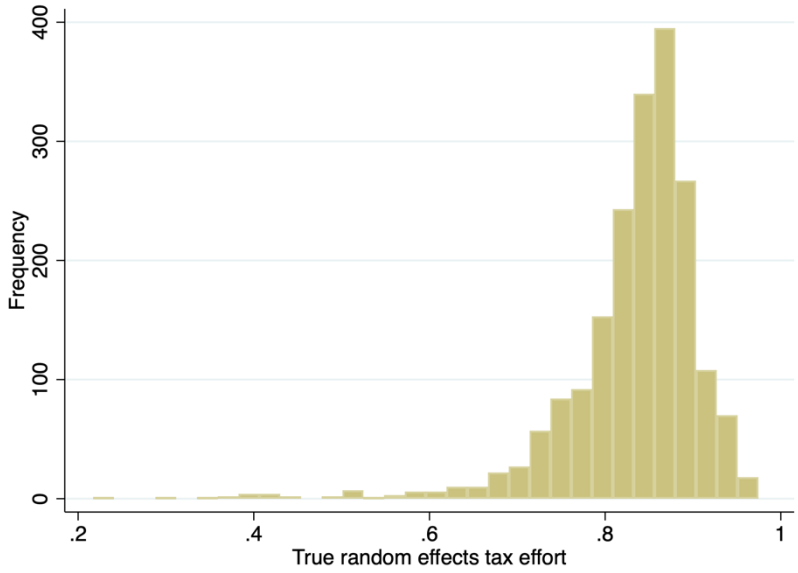
Source: own elaboration.

As it is observed from both frequency distribution graphs and the table of summary statistics both true random effects produced by McNabb et al. (2021) and those created by me in this paper follow a very similar distribution with few key differences. The minimum tax effort score of my estimates is much lower because the sample is reduced, and the standard deviation is larger for the same reason. The means and maximum values are comparable across the two samples.

Nevertheless, the estimates produced with baseline and augmented models resulted in changes in the summary statistics values. There is a slightly lower mean, lower minimum observed value, and the same maximum values for the augmented model when compared to the baseline. The standard deviation is also higher for the augmented model. Therefore, despite the exclusion of variables, the key difference comes from the sample selection. The use of additional variables does not have a significant influence on the tax effort scores observed in a country.

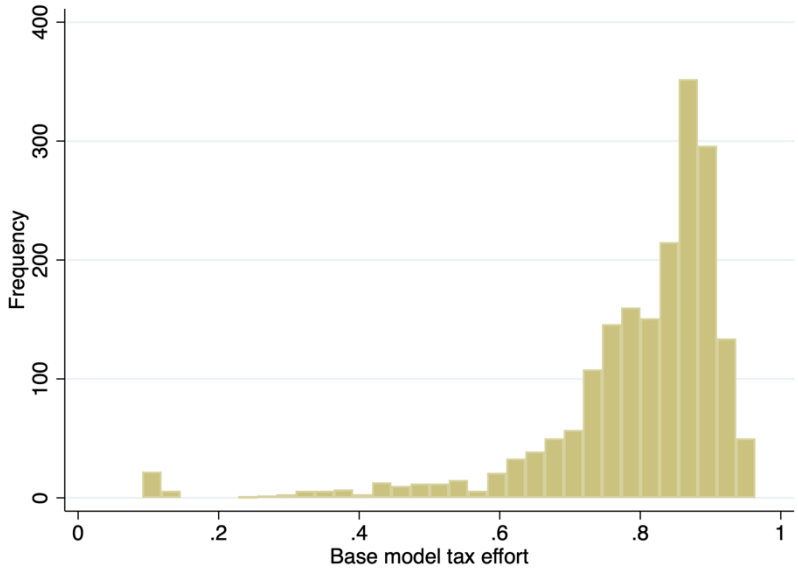
Most importantly, it is shown by comparing *Figures 14-16* that the distribution of tax effort does not change between the baseline in two papers and between proposed extended models. Therefore, estimates stemming from the extended frontier produced in this paper are well replicated, and the inclusion of additional variables has no substantial effects on the distribution of these scores. Altogether this makes reported tax effort scores better estimates of the true values as well as have those superior qualities of true random effect estimates as argued by McNabb et al. (2021).

Figure 14: Frequency distribution of tax effort scores in McNabb et al. (2021).



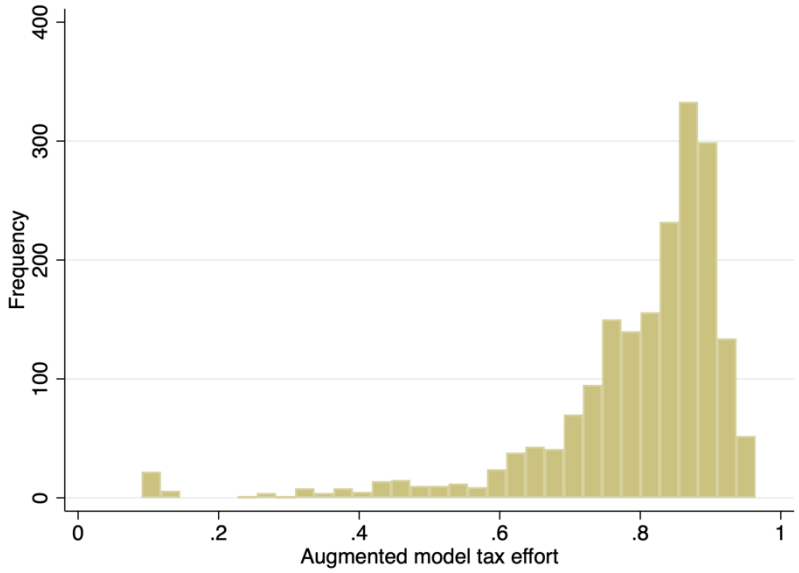
Source: own elaboration.

Figure 15: Frequency distribution of baseline tax effort scores.



Source: own elaboration.

Figure 16: Frequency distribution of augmented tax effort scores.



Source: own elaboration.

7. Conclusions and Discussion

This study was done to highlight shortcomings in the current literature on tax effort scores in both variables used and potential challenges. Despite very sophisticated methods of estimations, there does not exist a uniquely right way to create those scores and the method as well as variables depend on the discretion of the author. The stochastic frontier with true random effects is a prominent way to conduct those estimations while still yet to come under serious scrutiny. This paper presents evidence that these estimates are subject to potential issues in terms of sample heterogeneity and variable selection.

The first finding is that the estimates provided by the frontier have a very strong underlying assumption that all countries are essentially identical in their ability to collect revenues, which was proven wrong before and is demonstrated through regression analysis. If observations are split into sub-samples by income group, original variables are observed with statistically significant coefficients and their associations change depending on the sample. This presents an issue for research focused on the comparison of countries but using these international tax effort estimates available in UNU-WIDER GRD, or other tax effort estimates such as those available in USAID CTD. Therefore, the proper way to measure the tax effort would be to conduct the stochastic frontier analysis including only peer group countries, or equivalently those which are potentially subject to the same tax frontier as opposed to all countries available or provide a fully-fledged model based on these characteristics. The simplest way to do it is by constructing the frontier using the countries with the same level of income or other development proxies. Additionally, there are observed sign reversals for some variables based on the sub-sample analysis which also presents a problem in estimations as the effects of variables are heterogeneous for each income group. In this way, the current literature overlooks the fundamental differences between countries. Quantifying the importance of economic variables depending on development level as well as the production of more realistic tax effort scores are possible avenues for further research on this topic.

The next finding is that grants received by the country are associated with the tax effort score. Although omitted, this is an important variable to be included in the frontier estimations.

Moreover, foreign direct investment inflows and shadow economy are found to have statistically negative effects on the frontier estimations with true random effects. As argued above these additional variables do not affect the explanatory power of those previously used nor the distribution of tax effort score. Nevertheless, the inclusion of statistically significant variables theoretically makes estimates match more closely with the real-life potential tax revenues. However, there is a substantial problem of data availability. These variables reduce the sample for which those estimates can be provided while having arguably marginal effects on the final tax effort score reported. Thus, it may be logical to omit some of the statistically important variables in order to increase the coverage at the expense of precision which is the global aim of the UNU-WIDER GDR. Nevertheless, if the reduction of observations is not as significant, the inclusion of variables proposed in this paper will result in more sophisticated and generally superior estimations.

However, this study is not without limitations. The simple regression used for the tax effort scores has two main issues. The first one is that the lagged tax effort is omitted from the estimations, which may create problems of biases. Although the main conclusions do not come from these regressions the effect of the variables discussed may be incorrect. However, inferences drawn should not present these issues due to the use of robust standard errors. The common way to solve this problem is the application of dynamic panel data models such as generalized methods of moment estimations. However, these methods are outside the scope of this study. Additionally, this regression analysis is primarily focused on the collinearity problem of tax effort score and should be cautiously interpreted. As demonstrated by frontier estimation some variables such as foreign direct investment turn out to be significant for potential revenue estimation but will not be relevant in the regression analysis. Thus, the conclusions cannot be drawn from regression analysis alone, and all the variables must be included as controls in the frontier estimation. Moreover, all the limitations of tax effort studies are applicable here. For example, the total taxes excluding social contributions are subject to country-specific factors such as the difference between total taxes and total government revenues (arising from non-tax payments such as royalties) or the choice of funding of social security (either via tax system or earmarked contributions) (see McNabb et al., 2021).

Finally, as discussed previously these variables are not extensive, especially since there is an ongoing discussion on those aimed to explain institutional quality. One potential source of the data on institutional quality is variables of the tax administration. These variables are discussed and studied by dedicated agencies such as the Tax Administration Diagnostic Assessment Tool (TADAT). It may be possible to relate some aspects of the tax administration to the low observed tax effort score or failures in tax collection. However, these variables are extremely sensitive, have almost no coverage, and are inaccessible to the public. Nevertheless, one potential development in the literature on the tax effort would be the inclusion of tax administration quality measures in estimations which will both provide better scores and highlight important areas of tax administration to be modernized with the goal of revenue mobilization.

8. References

Alonso, J. A., & Garcimartín, C. (2011). *Does aid hinder tax efforts? More evidence* (CREDIT Research Paper 11/04). The University of Nottingham, Centre for Research in Economic Development and International Trade (CREDIT). https://hdl.handle.net/10419/65466

Ángeles Castro, G., & Ramírez Camarillo, D. B. (2014). Determinants of tax revenue in OECD countries over the period 2001–2011. *Contaduría y Administración*, *59*(3), 35–59. https://doi.org/10.1016/S0186-1042(14)71265-3

Angeles-Castro, G. (2006). The Effects of Economic Liberalization on Income Distribution: A Panel-Data Analysis. In E. Hein, A. Heise, & A. Truger (Eds.), *Wages, Employment, Distribution and Growth* (pp. 151–180). Palgrave Macmillan UK. https://doi.org/10.1057/9780230371781_8

Awasthi, R., & Engelschalk, M. (2018). *Taxation and the Shadow Economy: How the Tax System Can Stimulate and Enforce the Formalization of Business Activities*. World Bank, Washington, DC. https://doi.org/10.1596/1813-9450-8391

Bahl, R. W. (1971). A Regression Approach to Tax Effort and Tax Ratio Analysis (Analyse de l'effort et de la pression fiscale par la methode de regression) (Un estudio del esfuerzo tributario y de la presion fiscal mediante el analisis de regresion). *Staff Papers - International Monetary Fund*, *18*(3), 570. https://doi.org/10.2307/3866315

Bird, R. M., Martinez-Vazquez, J., & Torgler, B. (2004). Societal Institutions and Tax Effort in Developing Countries. *SSRN Electronic Journal*, *15*(1). https://doi.org/10.2139/ssrn.662081

Cassou, S. P. (1997). The link between tax rates and foreign direct investment. *Applied Economics*, *29*(10), 1295–1301. https://doi.org/10.1080/00036849700000019

Cyan, M., Martinez-Vazquez, J., & Vulovic, V. (2013). *Measuring tax effort: Does the estimation approach matter and should effort be linked to expenditure goals?* (ICEPP Working Papers 39). International Center for Public Policy. https://scholarworks.gsu.edu/icepp/39

Easterly, W., & Schmidt-Hebbel, K. (1991). *The macroeconomics of public sector deficits: A synthesis* (Policy Research Working Paper Series 775). The World Bank.

Eltony, N. M. (2002). *Determinants of Tax Efforts in Arab Countries* [Arab Planning Institute Working Paper 207]. Arab Planning Institute.

Fenochietto, R., & Pessino, C. (2013). Understanding Countries' Tax Effort. *IMF Working Papers*, *13*(244), 1. https://doi.org/10.5089/9781484301272.001

Gaspar, V., Jaramillo, L., & Wingender, P. (2016). *Tax Capacity and Growth: Is there a Tipping Point?* International Monetary Fund.

Ghura, D. (1998). Tax Revenue in Sub-Saharan Africa: Effects of Economic Policies and Corruption. *IMF Working Papers*, *98*(135), 1. https://doi.org/10.5089/9781451855685.001

Gupta, A. S. (2007). *Determinants of Tax Revenue Efforts in Developing Countries*. International Monetary Fund.

Hemming, R., & Kay, J. A. (1980). The Laffer Curve. *Fiscal Studies*, *1*(2), 83–90. https://doi.org/10.1111/j.1475-5890.1980.tb00554.x

Hinrichs, H. H. (1965). Determinants of Government Revenue Shares Among Less-Developed Countries. *The Economic Journal*, *75*(299), 546. https://doi.org/10.2307/2228952

IMF. (1971). Trends in taxation in developing countries. *Staff Papers - International Monetary Fund*, *18*(2).

IMF. (1975). Tax ratios and tax effort in developing countries, 1969–71. *Staff Papers - International Monetary Fund*, *22*(1).

IMF. (1979). International Comparisons of Taxation for Selected Developing Countries, 1972-76. *Staff Papers - International Monetary Fund*, *26*(1), 123. https://doi.org/10.2307/3866567

Kumbhakar, S. C., & Lovell, C. A. K. (2000). *Stochastic Frontier Analysis* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139174411

Langford, B., & Ohlenburg, T. (2016). *Tax revenue potential and effort: An empirical investigation. IGC Working Paper S-43202-UGA-1*. International Growth Centre. https://www.theigc.org/wp-content/uploads/2016/01/Langford-Ohlenburg-2016-Working- paper.pdf

Leuthold, J. H. (1991). Tax shares in developing economies A panel study. *Journal of Development Economics*, *35*(1), 173–185. https://doi.org/10.1016/0304-3878(91)90072-4

Lotz, J. R., & Morss, E. R. (1970). A Theory of Tax Level Determinants for Developing Countries. *Economic Development and Cultural Change*, *18*(3), 328–341. https://doi.org/10.1086/450436

Lutz, M., & Tanzi, V. (1991). *Interest Rates and Government Debt: Are the Linkages Global Rather Than National?* (IMF Working Papers 1991/006). International Monetary Fund. https://EconPapers.repec.org/RePEc:imf:imfwpa:1991/006

Mawejje, J., & Sebudde, R. K. (2019). Tax revenue potential and effort: Worldwide estimates using a new dataset. *Economic Analysis and Policy*, *63*, 119–129. https://doi.org/10.1016/j.eap.2019.05.005

McNabb, K., Danquah, M., Tagem, A. M. E., & UNU-WIDER. (2021). *Tax effort revisited: New estimates from the Government Revenue Dataset* (WIDER Working Paper 2021; WIDER Working Paper, Vol. 2021). UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2021/110-5

Medina, L., & Schneider, F. (2018). *Shadow Economies Around the World: What Did We Learn Over the Last 20 Years?* International Monetary Fund.

Michael Coppedge et al. (2024). *V-Dem Dataset v14* (Version 14) [dataset]. University of Gothenburg. https://doi.org/10.23696/MCWT-FR58

OECD. (2024). *Country statistical profiles: Key tables from OECD* [dataset]. OECD Publishing.

Olivera, J. H. G. (1967). Money, prices and fiscal lags: A note on the dynamics of inflation. *BNL Quarterly Review*, *20*(82), 258–267.

Oshima, H. T. (1957). Share of Government in Gross National Product for Various Countries. *The American Economic Review*, *47*(3), 381–390.

Pessino, C., & Fenochietto, R. (2010). Determining Countries' Tax Effort. *Hacienda Pública Española*, *195*, 65–87.

Pivovarsky, A., Clements, B. J., Gupta, S., & Tiongson, E. (2003). *Foreign Aid and Revenue Response: Does the Composition of Aid Matter?* International Monetary Fund.

Rogers, W. (1994). Regression standard errors in clustered samples. *Stata Technical Bulletin*, *3*(13). https://EconPapers.repec.org/RePEc:tsj:stbull:y:1994:v:3:i:13:sg17

Russell, B. (2010). *Revenue administration: Managing the shadow economy*. International Monetary Fund, Fiscal Affairs Dept.

Stotsky, J. G., & WoldeMariam, A. (1997). *Tax Effort in Sub-Saharan Africa*. International Monetary Fund.

Tanzi, V. (1977). Inflation, Lags in Collection, and the Real Value of Tax Revenue (L'inflation, les retards de recouvrement et la valeur reelle des recettes fiscales) (Inflacion, desfases en la recaudacion y valor real de los ingresos tributarios). *Staff Papers - International Monetary Fund*, *24*(1), 154–167. https://doi.org/10.2307/3866540

Tanzi, V. (1989). *The Impact of Macroeconomic Policies on the Level of Taxation and the Fiscal Balance in Developing Countries* (Vol. 36). International Monetary Fund.

Tanzi, V. (1992). Structural factors and tax revenue in developing countries: A decade of evidence. In I. Goldin & L. A. Winters (Eds.), *Open Economies* (1st ed., pp. 267–285). Cambridge University Press. https://doi.org/10.1017/CBO9780511628559.023

UNU-WIDER. (2023). *UNU-WIDER Government Revenue Dataset – Version 2023* (p. Version 2023) [dataset]. https://doi.org/10.35188/UNU-WIDER/GRD-2023

Williamson, J. G. (1961). Public Expenditure and Revenue: An International Comparison. *The Manchester School*, *29*(1), 43–56. https://doi.org/10.1111/j.1467-9957.1961.tb01187.x

World Bank. (n.d.-a). *How does the World Bank classify countries?* The World Bank. Retrieved May 26, 2024, from https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries

World Bank. (n.d.-b). *Taxes & Government Revenue*. The World Bank. Retrieved May 26, 2024, from https://www.worldbank.org/en/topic/taxes-and-government-revenue

World Bank. (2023). *World development indicators* [dataset]. The World Bank Washington DC. https://databank.worldbank.org/source/world-development-indicators

9. Appendix I

Table 8: Appendix I: Fixed effects regression with robust standard errors for tax effort for all variables.

	(1) Tax effort (TREz)
GDP per capita, PPP (current international $)	0.00000184
	(0.55)
GDP per capita, PPP (current international $) squared	-8.77e-12
	(-0.22)
Trade (%GDP)	-0.000799**
	(-2.26)
Total Resource Revenue	0.0260***
	(3.95)
Agriculture, forestry, and fishing, value added (% of GDP)	-0.000417
	(-0.16)
Particularistic or public goods	-0.0436**
	(-2.20)
Urban population (% of total population)	-0.00543*
	(-1.82)
Public sector corruption index	-0.474**
	(-2.41)
Accountability index	0.145*
	(1.73)
Rule of law index	-0.757**
	(-2.17)
Shadow economy index	0.00313
	(1.13)
Grants	0.0347***
	(2.81)
Foreign direct investment, net inflows (% of GDP)	-0.0000212
	(-0.05)
Inflation, consumer prices (annual %)	0.0000718
	(0.10)
Gross fixed capital formation	0.000584*
	(1.73)
Central government debt, total (% of GDP)	0.0000508
	(0.20)
Constant	1.772***
	(6.20)
Observations	394
Adj. R-squared	0.321

Note: *t* statistics in parentheses and based on robust S.E.; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; *Source: own elaboration*