Do Tax Audits Have a Dynamic Impact? Evidence from Corporate Income Tax Administrative Data

Christos Kotsogiannis[a], Luca Salvadori[b,†], John Karangwa[c], Theonille Mukamana[d]

First version: September 7, 2021
This version: March 25, 2024

Abstract: Making use of a unique administrative data set consisting of the universe of administrative filings in Rwanda, this paper investigates the impact of tax audits on businesses' reporting behaviour. The evidence suggests that tax audits have a positive impact on corporate income and corporate tax liabilities reported for three years after the start of the audit process. The results also suggest that the type of audit matters. While 'comprehensive' tax audits have a significant positive effect on compliance, 'narrow-scope' tax audits exhibit both a positive and a negative effect during a three-year period after the audit, with the net impact being negative. The implication of this, from a tax compliance perspective, is that 'narrow-scope' audits are ineffective and that doing more of those and less of comprehensive ones might have a negative impact on tax compliance. Effective tax compliance strategy therefore requires the careful evaluation of all types of audits.

Keywords: Tax Audit Evaluation; Tax Administration; Tax Evasion; Tax Compliance.

JEL classification: H25, H26, H32.

---

[a] Tax Administration Research Center (TARC), Department of Economics, University of Exeter Business School and CESIfo.

[b] Autonomous University of Barcelona (UAB), Barcelona School of Economics (BSE), Tax Administration Research Center - TARC (University of Exeter) and Barcelona Institute of Economics (IEB).

[c] Planning and Research Department, Rwanda Revenue Authority, Kigali, Rwanda.

[d] The African Institute for Mathematical Sciences Research and Innovation Centre, Lappeenranta-Lahti University of Technology, and Planning and Research Department, Rwanda Revenue Authority at the time of involvement in this project.

[†] Corresponding author. Email: luca.salvadori@uab.cat.

1 Introduction

Recent estimates have it that achieving the Millennium Development Goals requires increasing domestic revenues in low-income countries by around 15 points of GDP, a target which requires the implementation of key policy reforms (Gaspar et al., 2019). While the effort to boost domestic revenue mobilization in developing countries continues, it is now even more challenging, and pressing, given the recent and much needed relief measures implemented to ease the impact of the COVID-19 pandemic.[1] Amongst the policy reforms for improving revenue mobilization is strengthening tax administration capacity, an issue which has come to the fore for many countries around the world during the last two decades.

An integral part of tax compliance is operational audits, and the extent to which they are conducive to future compliance. There is a growing literature on audit assessment providing mixed evidence regarding the impact of tax audits on future compliance.[2] Surprisingly, however, little attention has been paid to the compliance impact of the types of audit, which are broadly categorised as 'comprehensive' and 'narrow-scope' (with the latter further categorised as desk-based or issue-oriented). Comprehensive audits are in-depth and in-person examinations conducted across different tax bases and fiscal years, whereas narrow-scope audits focus on a limited number of fields in a tax return. This distinction in audit types, and their impact on tax compliance, is at the heart of the contribution of this paper.

More specifically, this paper investigates the impact of auditing on deterring future noncompliance of incorporated businesses in Rwanda, paying particular attention to evaluating the impact of the different types of tax audits performed by the Rwandan Revenue Authority (RRA). Focusing on Corporate Income Tax (CIT) audits is important as reliable evidence is lacking and potentially large sums of underreported revenues are involved

[1] Between 2010 and 2019 the average tax-to-GDP ratio in Africa (30) increased by 1.8 percentage points, an increase which is similar in magnitude to the increases in Latin American Countries (LAC) and the OECD averages during the same period. Non-tax revenues, however, decreased substantially and by around 1.8 percent of GDP. During this period external debt costs also increased by 1.1 percent of GDP a figure that is expected to rise considerably as a consequence of the adverse impact of the COVID-19 pandemic, OECD et al. (2021).

[2] As an illustrative list of contributions (relying on data from high-income countries) see Kleven et al. (2011), Løyland and Øvrum (2017), DeBacker et al. (2018a,b), Beer et al. (2020), and Advani et al. (2021) presenting evidence for a positive impact of audits, and Kasper and Rablen (2023), Alm and Malézieux (2021) and Antinuan and Asatryan (2020) for negative. This latter possibility has been until recently only theoretical. This effect has been mostly associated with what colloquially has been described as the 'bomb-crater' effect, and driven by the perception of the taxpayer that lightning (a metaphor for audit) does not strike twice. Along the same line Mendoza et al. (2017), using country-level data, report evidence of a U-shaped relationship between the level of tax auditing and tax evasion supporting somewhat the notion that fiscal controls may lead to an increase in noncompliance.

(Slemrod, 2019).[3] Focusing on Rwanda presents a unique opportunity in understanding how well tax audits perform in a developing country whose economy over the last decade has been growing steadily. Indeed, research has recently begun to pay attention to developing countries, and there is now an emerging literature analysing tax compliance issues from a number of different perspectives.[4] But the evidence on the compliance impact of tax audits is still rather limited, considering their importance for much needed domestic revenue mobilisation.[5]

To address the issue discussed above use is made of unique administrative data consisting of the universe of incorporated businesses' anonymized tax declarations, the universe of risk-based anonymized audit data for the audit wave in 2015, as well as detailed information on the risk rules employed by RRA to prioritise CIT audits. Tax audits are prioritized by RRA through a risk-based assessment procedure and therefore the audited taxpayers might not be (in a statistical sense) identical to the unaudited ones. To address this potential selection bias the analysis employs a matched-Difference-in-Difference approach (matched-DID).

The results provide evidence of a significant and increasing *pro-deterrence* effect of audits on corporate taxable income and CIT liability reported over the three years after the audit process.[6] The estimates reveal that the total dynamic deterrence effect associated to this change in reporting behaviour amounts to approximately 16.6 percent of the total revenues collected through the audit process, which include also the detected underreporting and penalties. This pro-deterrence effect is driven by comprehensive and not narrow-scope audits (desk-based and issue-oriented); with the latter having a positive pro-deterrence impact only one year after the audit, and turning to counter deterrence from then on effectively outweighing the pro-deterrence impact observed in the first year.[7] A possible explanation for this behaviour is that a taxpayer who experiences

[3]Generally, the literature has focused on individual tax behaviour with notable exceptions being DeBacker et al. (2015) and Li et al. (2019) which investigate corporate tax behaviour.

[4]See, for example, Brockmeyer et al., 2019 on the role of communication in compliance in Costa Rica, Waseem (2021) on the role of withholding in the self-enforcement implicit mechanism of VAT in Pakistan, Balán et al. (2021) on the role of tax collectors in property taxes, Bergeron et al. (2020) on property taxes and the Laffer curve in D.R Congo, Tourek and Dada (2021) on the role of peer information and social norms for compliance in Rwanda. See also Mascagni et al. (2016, 2022) who investigate the nexus between tax compliance and progressivity and the phenomenon of nil-filing in Rwanda. Ebrahim et al. (2021) evaluate a pilot program introducing a risk-based system for audit selection in Tanzania showing that the intervention increases adjusted taxable income by about 15 percent on the first year of implementation and Best et al., 2021 on the deterrence value of VAT audits in Pakistan.

[5]Perhaps the neglect comes from the fact that quality tax administrative data has been until recently difficult to come by.

[6]Taxable income reported and CIT liability reported are of course related, but given the existence of different tax regimes their relationship is not one-to-one. For this reason the analysis reports estimates on both outcomes. See also Section 4.

[7]The results obtained on audit type relate, conceptually, to a recent report utilising U.S. data finding

a more mechanical audit,[8] and have underreported on fields not examined under a desk audit, may revise their expected gain from noncompliance upwards thereby increasing noncompliance, following the audit. The implication of this, from a tax compliance perspective, is that 'narrow-scope' audits are ineffective and that doing more of those and less of comprehensive ones might have a negative impact on tax compliance. This does not of course suggest that narrow-scope audits are not a desirable instrument for compliance. What the analysis here points to is that the presumption that if Revenue Authorities do more of those and less of comprehensive tax compliance will increase might be incorrect: Effective tax compliance strategy requires the careful evaluation of all types of audits.

The reminder of the paper is organized as follows. Section 2 reviews the literature placing the paper within the broader scholarly research area of tax audits evaluation, and Section 3 provides a conceptual framework whose sole purpose is to explore the role of information provided by audits in taxpayers' compliance behaviour and thus rationalize the empirical results derived. Section 4 presents the institutional setting and the data the analysis is based on. Section 5 describes the methodological approach followed, and Section 6 presents the results. Section 7 provides some concluding remarks.

2 Literature review

This paper contributes to several strands of the literature that evaluates the impact of tax audits on audited taxpayers.[9] First, it contributes to the emerging literature that evaluates the impact of audits that differ in their intensity and scope. As noted in the introductory section, like many countries around the world, RRA, and in addition to comprehensive audits, relies also on audits that are narrower in their scope. Research on this is scant, with notable exceptions being the empirical contribution of Erard et al. (2019) and the contribution of Kasper and Alm (2022) who provide experimental evidence that tax audits may have differential effects on compliance. Like them, this paper finds support that these types of audits have the opposite impact to the one intended for: comprehensive audits increase compliance but narrow focused do not.

The paper also contributes to the evidence basis of the deterrence effect of audits on corporate tax behaviour. To the best of our knowledge there are only two contributions

that correspondence audits are not a perfect substitute for face-to-face examinations with the former being generally associated with a counter-deterrent effect while the latter with a pro-deterrence effect (Erard et al., 2019).

[8]And on a single issue, such as VAT refund, declared expenses for tax deductions or insufficient documents to assess tax liability. This presumes that the finding of the audit neither triggers a comprehensive one (or subsequent narrow-scope ones) nor it penalises heavily the taxpayer.

[9]For a more elaborate review across the many dimensions of audits, see Kotsogiannis and Salvadori (2024).

that evaluate the future impact of audits on corporate behaviour using business level administrative data.[10] Li et al. (2019) find evidence for a pro-deterrence effect of corporate tax audits on businesses' behaviour using data obtained from a local tax office in China, showing that after firms have been audited they significantly increase taxes paid, reduce their book-tax differences, and also reduce their income-decreasing discretionary accruals. Tax audits might also increase corporate tax aggressiveness, a finding reported in DeBacker et al. (2015) who, using U.S. data, provide evidence of an increase in corporations' tax aggressiveness for a few years after having received an audit and show that tax aggressiveness progressively reduces with time.

Lediga et al. (2020) investigate corporate tax audits in South Africa, focusing on the extent to which audits have an impact on neighboring (that is, those that are in close proximity) businesses, showing that the impact is short-lived and levels off two years after the audit has taken place. Focusing on VAT, Best et al. (2021) exploit a national program of randomized VAT audits in Pakistan to investigate how much evasion audit uncovers and how much evasion it prevents through changing behaviour of businesses, finding that although tax audits uncover a substantial amount of evasion they do not deter future compliance. Best et al. (2021), based on interviews they had with tax auditors, rationalise their finding by suggesting that audits in Pakistan tend to focus on checking mechanical violations of law which typically are likely to result in additional revenue but unlikely to move firm priors on the detection probability outward.

Like most of the recent contributions employing different methodologies across different contexts,[11] this paper finds a significant positive impact of tax audits on the future reporting behaviour of audited taxpayers. Interestingly, this is not the case for narrow-scope audits, with the evidence suggesting a reverse of the effect after the first post-treatment year leading to a net negative impact overall.

3 Conceptual framework

This section provides a conceptual framework whose sole purpose is to describe a mechanism through which tax audits might affect taxpayers' future compliance behaviour. There are two elements underlying this mechanism. First, audits performed by RRA *partially* reveal the extent of underreporting, either because audits are narrower in their scope (and less intensive) or because (even if they are comprehensive) they do not detect the true extent of underreporting. Secondly, taxpayers rationally utilise the available information,

[10] Additional contributions—but based on indirect approaches and with mixed evidence—are Atwood et al. (2012), Hoopes et al. (2012), Finley (2019), and Eberhartinger et al. (2020).

[11] See, for example, Kleven et al. (2011), Løyland and Øvrum (2017), DeBacker et al. (2018a), DeBacker et al. (2018b), Beer et al. (2020), and Advani et al. (2021).

which can be obtained from various sources, including the audit process and its outcome, in order to infer the true probability of them being audited. Both of these elements put together affect the trade-off between underreporting and being caught (and penalised) and not being caught and, therefore, the level of underreporting given true income.[12] It is conceivable, for example, that if tax audits provide the wrong type of information to the taxpayers (relative to some initially held belief regarding their likelihood of them being audited and found noncompliant), they might rationally decide to become less compliant in the subsequent years. The discussion next turns to further elaborating on this mechanism .

3.1 On the role of information

To elaborate on the role of information obtained in taxpayers' updating their beliefs regarding the true likelihood of auditing, suppose that at time τ the taxpayer has access to a prior distribution, denoted by $g(p)$, which gives the probability, denoted by p, of them being audited (and found noncompliant) with its mean and variance being given by $E(p)$ and $Var(p)$, respectively. This prior might reflect past experience and/or information provided by RRA including information transmitted through their business plans but also the annual reports regarding the (aggregate) likelihood of auditing.[13] The prior belief of taxpayers might not reflect precisely the true probability distribution of p, an aspect that is captured by the $1/Var(p)$. Denote the information received by the taxpayer[14] by $\tilde{p} \in [0, 1]$ and assume that this information is unbiased in the sense that conditional on the true probability of auditing the expected value of the information received is the true likelihood of auditing, that is $E(\tilde{p}|p) = p$. The information transmitted is unbiased but not precise an aspect that is captured by the $1/E(Var(\tilde{p}|p))$, which gives the precision of the information received (if $1/E\left(Var(\tilde{p}|p)\right) \rightarrow 0$ (∞) then the information is inaccurate (accurate)).

Given the information received, businesses update their prior belief. Assuming that the posterior density $z(\tilde{p}|p)$ and the prior density $g(p)$ give rise to a linear posterior density,

[12] The reasoning developed here follows that of Advani et al. (2021), DeBacker et al. (2015), and Best et al. (2021).

[13] The RRA, as other Revenue Authorities do, does publish the total number of audits to be conducted throughout the year in its business plan and also (though less often) the sectors which maybe targeted through the audit campaign. In the annual reports they, too, publish aggregate information on the performance of these audits. Nevertheless, since this information is neither business specific nor the exact criteria used in the risk-based assessment are known, business cannot infer accurately the likelihood of them being audited.

[14] The information received could be called 'signal' which is received by the taxpayer and rationally utilized to infer the true value of the audit probability.

the expected probability of being audited[15] at time τ is given by

$$E(p|\tilde{p}) = \left(\frac{\frac{1}{Var(p)}}{\frac{1}{Var(p)} + \frac{1}{E(Var(\tilde{p}|p))}} \right) E(p) + \left(\frac{\frac{1}{E(Var(\tilde{p}|p))}}{\frac{1}{Var(p)} + \frac{1}{E(Var(\tilde{p}|p))}} \right) \tilde{p}\,, \tag{1}$$

and so it is a *weighted average*[16] of the business/taxpayer's prior mean of the probability of being audited $E(p)$ and the information obtained from the audit $\tilde{p}$, at time τ, with the weights depending on the precision of the prior distribution, $1/Var(p)$, and the precision of the information obtained, $1/E\left(Var(\tilde{p}|p)\right)$. Differentiating (1) with respect to $1/E\left(Var(\tilde{p}|p)\right)$ gives

$$\frac{\partial E(p|\tilde{p})}{\partial \left(\frac{1}{E(Var(\tilde{p}|p))} \right)} = - \frac{\frac{1}{Var(p)}}{\left(\frac{1}{Var(p)} + \frac{1}{E(Var(\tilde{p}|p))} \right)^2} \left(E(p) - \tilde{p} \right)\ , \tag{2}$$

and so the expected probability of auditing $E(p|\tilde{p})$ is decreasing in the precision of the information received by the businesses, $1/E\left(Var(\tilde{p}|p)\right)$, if and only if the expected prior, $E(p)$, is greater than the information received, $\tilde{p}$; otherwise it is increasing.

3.2 Likelihood of auditing

The point thus far—and central to the analysis in this paper—is that information matters for the taxpayers and it informs the estimated probability of auditing which affects their decision to underreport.[17] If audits are not informative for the businesses then there is no reason for them to change their future compliance behaviour, for given true income. If, on the other hand, current audits convey accurate, and therefore valuable, information for the businesses then a businesses will rationally incorporate this in their decision to underreport in the future. It is therefore conceivable that audits (to be interpreted broadly) might have a negative impact on compliance. This will be the case if current audits are more uninformative from the previous ones: In this case the taxpayers will rationally rely more on their prior regarding the audit probability. This follows directly from (1): for given $\tilde{p}$, (1) implies that $E(p|\tilde{p})$ is decreasing in the precision of the information obtained if and only if $E(p) > \tilde{p}$, and increasing otherwise. This is the case, if the expected prior mean of the taxpayer is higher than the information obtained (and this information is accurate)

[15]This follows from Erikson (1969). For an application of this in taxation matters see Kotsogiannis and Serfes (2014).

[16]Notice that equation (1) is satisfied under prior-posterior distribution functions, Beta-Binomial, Gamma-Poisson and Normal-Normal.

[17]Implicit in this discussion there is a time dimension: information received at time τ informs the optimal decision to underreport of the taxpayer at time $\tau + 1$. Also the RRA, consistent with Allingham and Sandmo (1972), is not an active player. This is, arguably, a limitation which, however, does not affect qualitatively the important role the information plays in compliance, as long as there is a capacity constraint on the part of RRA.

then the taxpayer revises their belief regarding the probability of being audited next period downwards. If underreporting is a decreasing function of the expected probability of being audited then underreporting will increase (and vice versa if $E(p) < \tilde{p}$).

3.3 Information and decision to under-report

To put the above into context consider the canonical model developed and analyzed by Allingham and Sandmo (1972) but appropriately modified to incorporate the expectation derived in equation (1). In this model a taxpayer decides whether, and how much, to evade their tax liabilities, a decision which is influenced by existing penalties (and the legal environment) if the taxpayer is audited and found underreporting their tax liabilities. Denote the true income of the taxpayer by y, the taxpayer who has verified income x pays a proportional tax, denoted by t on declared income, and so tx. The taxpayer is also aware that if they are audited their true income y will be determined with certainty, and they will have to pay all additional taxes due plus a penalty. If income underreported is discovered (in the sense that income declared x is less than true income y) through auditing (which occurs with probability p), the taxpayer pays a penalty, denoted by π, proportional to the amount of income underreported that is, $\pi t(y-x)$. The total amount the taxpayer pays in this case is $ty+\pi t(y-x)$ and the realised income is $Z = y(1-t) - \pi t(y-x)$. If on the other hand the taxpayer is not audited their true income is given by $Y = y - tx$. Recalling that the expected probability of auditing is given by (1), and assuming risk aversion,[18] the taxpayers maximise expected utility, denoted by W and given by,

$$\max W = E\left(p|\tilde{p}\right)U(Z) + \left(1 - E(p|\tilde{p})\right)U(Y), \tag{3}$$

by choosing how much income x to report to the Revenue Authority, with the optimal $x\left(y, E\left(p|\tilde{p}\right), \pi, t\right)$, and so noncompliance $y - x(y)$, being determined by the necessary condition (for an interior solution)[19]

$$W_x\left(x; y, E\left(p|\tilde{p}, \pi, t\right)\right) = (1 - E\left(p|\tilde{p}\right))U_Y(Y) - E\left(p|\tilde{p}\right)U_Z(Z)\pi = 0. \tag{4}$$

It is straightforward to show, following Allingham and Sandmo (1972), that an increase in the (expected) probability of auditing $E\left(p|\tilde{p}\right)$ reduces underreporting in the sense that $x_{E(\cdot)} < 0$, since it makes the act of underreporting more expensive for the taxpayer. The question now is how does the information content of audits affect compliance? The

[18] A subscript denotes differentiation with respect to the argument given.

[19] Notice, as already noted above, the audit probability in principle can be conditioned on the income reported x, as in Reinganum and Wilde (1986). In addition, audit success on the part of the Revenue Authority depends on the intensity and quality of an audit which is all subsumed within the function $p(x)$, see Kotsogiannis and Serfes (2016).

answer to this relies on how the expected probability of auditing, $E(p|\tilde{p})$, is affected by the information obtained by the taxpayer. Following (2), if the information obtained from audits is uninformative (in the sense that $1/E(Var(\tilde{p}|p))$ is significantly low with the extreme being 0) then the taxpayer learns nothing from audits and therefore it is rational that they put more weight on the mean of the prior distribution of auditing. In this case the taxpayer, for given income y, chooses the level of x basing their decision more on the prior mean $E(p)$. If, on the other hand, audits convey information, in the sense of $1/E(Var(\tilde{p}|p))$ being significantly high, then tax audits are informative and so more weight in updating the beliefs regarding the probability of auditing is put on the information obtained.

The framework outlined is not indented to model the behavioural response of the taxpayer when faced with the two types of audits, but its purpose is to emphasise that the accuracy of an audit matters for the underreporting decision of the taxpayer. Whether comprehensive tax audits are conducive to more or less compliance than narrow-scope will of course depend on a number of factors, such as the level of fine under each audit, the extent to which narrow-scope audits might trigger a comprehensive one, as well as the impact on the taxpayer's risk profile which might influence future risk assessment (and so the frequency of the audit) of the taxpayer. Narrow-scope audits are typically more mechanical and therefore easier to be manipulated by the taxpayer.[20] As such if narrow-scope audits identify less underreporting than the true level (and relative to the comprehensive ones) then, ceteris paribus, one would expect that they are conducive to less compliance.

The next section presents the institutional setting and the data used in the analysis.

4 Institutional setting and data

Rwanda is a representative low-income country, with a tax-to-GDP ratio similar—in both level and trend—to the average of African peer countries and increasing, and therefore provides an interesting framework for assessing tax audit strategies in developing countries.[21]

RRA classifies businesses as follows: *Micro*-businesses are defined as those declaring a turnover of less than 12 million Rwf (USD 13,380 as of February 2019 exchange rate) in

[20] Much like as in Best et al. (2021).

[21] The tax-to-GDP ratio in Rwanda has been steadily increasing from 10 percent in 1998 to 16.3 percent in 2018-2019, RRA (2019). In terms of structure, tax revenues follow the average of a low-income country with reliance on CIT and VAT which collect almost half of the total tax revenues (46 percent), RRA (2019), lower than the 48 percent average of Africa, and higher than Latin America and Caribbean lower (44 percent); while the average share is significantly lower in OECD countries (29 percent).

a tax period; *Small*-businesses have a turnover between Rwf 12 million and Rwf 50 million (USD 55,750); *Medium* and *Large*-businesses have a turnover higher than Rwf 50 million. Any person/business subject to any type of tax administered by RRA has to be registered with RRA and obtain a fiscal number before engaging in any economic activity. Rwanda collects around 50 percent of its tax revenues from CIT (the average of the CIT tax base for 2013-2018 is 17.24 percent) and VAT (the average of the VAT tax base for 2013-2018 is 33.06 percent). The CIT is a tax on income generated by incorporated businesses, and has to be declared and paid annually before April (by 31st March) of the following tax period. The are three CIT regimes: The CIT-real, CIT-lump sum and CIT-flat tax.

The CIT-real regime entails a standard tax rate of 30 percent on profit with some reductions available for specific groups. Small businesses can decide to benefit from a simplified CIT-lump sum tax regime having to pay a lump sum tax at the rate of 3 percent on their turnover while micro-businesses companies pay a CIT-flat-tax between Rwf 60,000 and Rwf 300,000, as classified by their turnover.

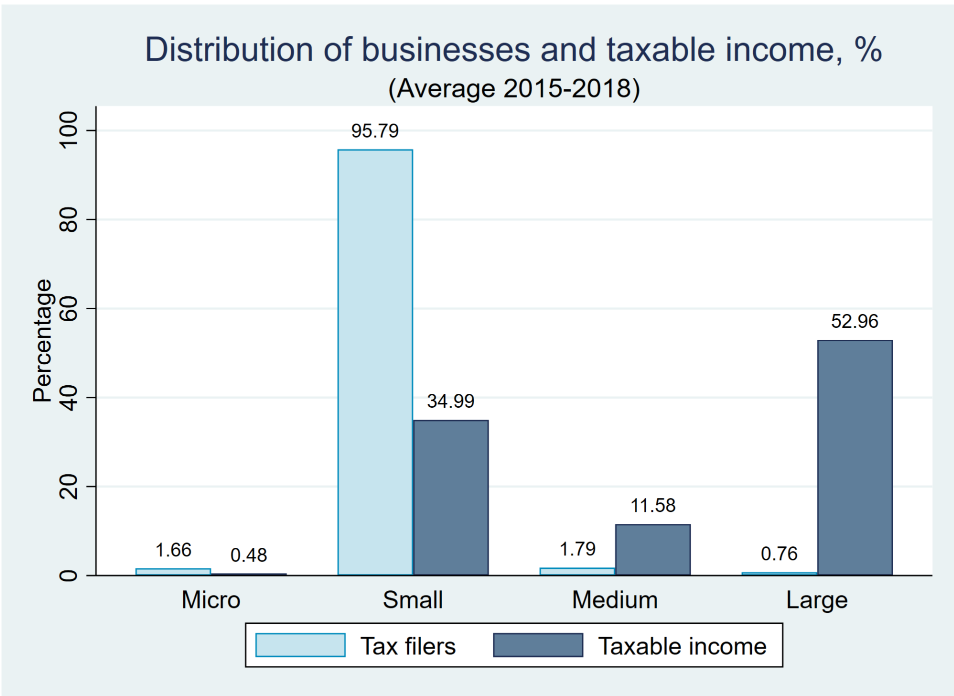
Businesses are required to file a CIT declaration form annually and irrespective of the CIT regime, and CIT can be prepaid in quarterly instalments. Businesses reporting under the CIT-real regime provide detailed information on the amount of business income and income from other sources, total expenses and depreciation income, and deductions, all of which determine the taxable income, as well as tax discounts and credits which define the tax liability owed by the business in that tax year, among other items. Businesses under the CIT-lump-sum regime are also required to file a CIT declaration annually but a significantly less detailed one. The information provided under the CIT lump-sum regime includes income from different sources—which determine the taxable income and tax liability—and withholding taxes. Businesses under the CIT-flat regime are required to file a considerably simplified form including their business income, which determines the amount to be paid, and tax credits claimed that coincide with the sum of quarterly instalments already paid.

The analysis focuses on taxable income and CIT liability reported by businesses as the two outcome variables. Taxable income reported across the tax regimes is the tax base upon which the corporate income tax is applied, while CIT liability is the tax payable by taxpayers net of any tax discounts claimed.[22]

[22]More precisely, for CIT-real regime taxable income corresponds to the total income obtained from different sources and is calculated net of expenses, depreciation adjustments and deductions. When this calculation leads to a negative amount, the taxable income reported is null as the business is not required to pay CIT and may carry forward the registered loss as a deduction in the following declarations (up to a period of five years). For businesses declaring under CIT-lump sum regime, taxable income reported corresponds to the total income declared from different sources while for taxpayers reporting under the CIT-flat regime taxable income coincides with their business income. Given their favourable tax schedule, business reporting under the two simplified regimes may not reduce the taxable income through deductions

All data employed in this paper is at the taxpayer (business) level:[23] They include mostly financial variables used to calculate taxes (for example, total sales, taxable income, VAT refunds), as well as some taxpayer characteristics, such as sector and geographical location (at tax centre level). Figure 1 presents the average distribution of businesses reporting taxable income by size across the period 2015-2018 together with the corresponding share of taxable income declared by size of business. The large majority of businesses filing a CIT declaration in Rwanda in this period are identified as Small-businesses (95.79 percent), with Medium-businesses and Large-businesses consisting together of roughly 2.55 percent of the total, while Micro-businesses represent 1.66 percent of the population. Not surprisingly, despite their small number, the main share of taxable income is reported by large/medium businesses, which account together for 64.54 percent of total taxable income declared (52.96 percent for large and 11.58 percent for medium businesses). This pattern is even amplified in terms of revenues collected. Figure 2 shows that the main share of tax revenues is collected from large/medium businesses, which account together for 81.94 percent of total revenues collected (68.54 percent for large, and 13.40 percent medium businesses).[24]

Figure 1: Distribution of businesses and taxable income (2015-2018)



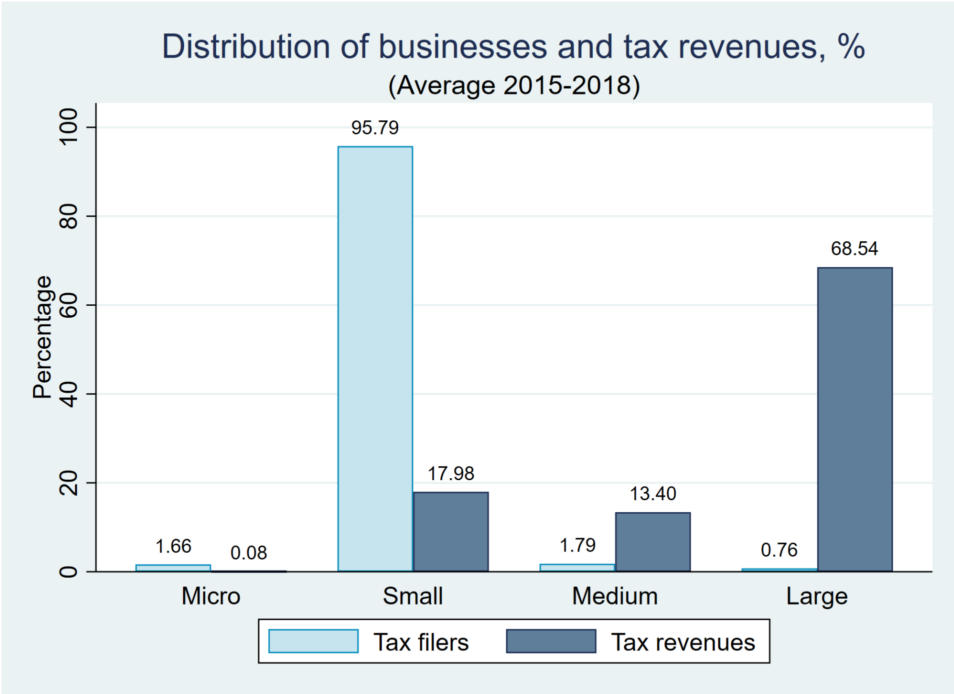
Note: Authors' calculations based on data provided by RRA.

nor CIT through discounts.

[23]See also Appendix A for additional information on the data set.

[24]Interestingly, as shown in Appendix A, there is a similar U-shaped relationship between audit probability and taxable income reported which also relates to the riskiness of noncompliance (see in particular Figure A.5 and A.7).

Figure 2: Distribution of businesses and revenues collected (2015-2018)



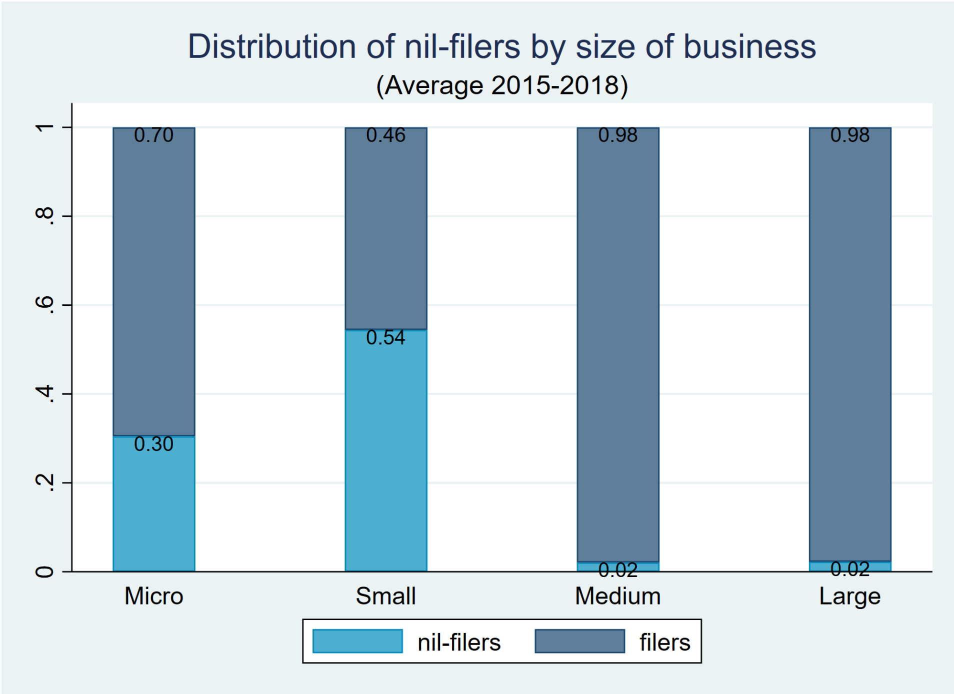
Note: Authors' calculations based on data provided by RRA.

A significant share of CIT filers (around 55 percent in 2015) are taxpayers who have not reported any business activity during a given tax period (known as 'nil-filers'). Figure 3 presents the average share of nil-filers and filers (those who submitted a tax return and declared positive taxable income) by size of business across the period 2015-2018. The majority of nil-filers belong to the categories of small and micro-businesses. In particular, 30 percent of micro-businesses are nil-filers, while 54 percent of small-businesses report zero on all fields of their tax declaration. The percentage of nil-filers is smaller for medium (2 percent) and large-businesses (2 percent).

To assess the impact of tax audits on CIT filers' reporting behaviour the analysis relies on different data sets provided by RRA: The world of (anonymised) CIT declarations for the tax periods from 2013 to 2018, as well as detailed records of audits undertaken by the Large Taxpayers Office, Small and Medium Taxpayers Office, and the Regions and Decentralised Tax Office during the years 2013 through 2016. We have also been given access to the tax audit appeals which is an important element of the audit process. Given the small number of appeals related to audits in the 2015 wave they do not though seem to play a crucial role in determining the results of the analysis.

The analysis will also utilise detailed confidential information on the criteria for audit selection which includes the risk rules employed to assign risk scores to the world of

Figure 3: Nil-filers and Filers by size (2015-2018)



Note: Authors' calculations based on data provided by RRA.

CIT declarations. The risk criteria utilise information that spans across tax bases.[25] The administrative data is retrieved from the RRA systems which collect and store tax data from tax procedures followed by taxpayers.

In general, the RRA tends to audit two tax periods but taxpayers are required to keep their records for a period of ten years. Tax enforcement examinations involve different types of audits:[26] Desk audits, issue audits (both of which are classified as narrow-scope) and comprehensive audits. Desk audits utilise information already submitted to RRA through various sources including from the declarations of many tax types, including VAT. These audits are conducted if the turnover of VAT does not correspond to the turnover of income tax without justification; if the tax declarations are not corresponding to paid taxes; if the taxpayer deducted from taxable income non-deductible expenses; if one or more invoices were not declared; in any other situation where the tax administration has sufficient documents that can be used to assess taxes. In a desk audit the taxpayer is not necessarily informed by the RRA, but it is invited for explanations before the tax

[25]After each return has been filed, audit flags are generated based on the characteristics of the returns in a deterministic way. In conducting the audits, tax auditors follow the audit procedures described in the manual of audits which provides a systematic approach to the tax audit process ensuring consistency in auditing. The integrity of the tax and audit data has been assured by the RRA.

[26]Following an administrative procedure RRA may also amend submitted tax liability which is initiated when the tax administration discovers a miscalculation or omission, an understatement or any other error in which case the tax administration rectifies the submitted tax liability. These amendments are not considered audits and therefore they do not appear in the analysis.

notification is issued. An issue audit usually focuses on a single tax type, single aspect or single tax period (for example, refund audits are a type of issue audit which focuses on tax declarations claiming refunds, VAT or income tax, from the RRA). Issue audits may be desk-based or, depending upon the nature of the inquiry, they may involve visits to the taxpayer's business premises. Comprehensive audits are more in-depth and time-intensive and usually are conducted through RRA staff visiting the taxpayer's business premises in order to review all relevant documents. During 2015 the RRA performed 435 audits involving corporate taxable income reported, with 389 of those (and so 89.43 percent of all audits) uncovering taxable income underreported by taxpayers, with 217 leading to the application of tax fines (which amounts to 49.89 percent of the total).

Table 1 presents summary statistics for the main outcome variables associated with the tax audits performed by the RRA in 2015 and for all tax periods audited. The maximum amount of taxable income underreported uncovered (given by *audit outcome*) is just over US$ 19,000,000, with the mean being just over US$ 101,000 and the standard deviation just under US$ 1,000,000. *Audit outcome* is also reported as a share of the *potential tax base* (defined as the sum of taxable income declared by the taxpayer and the *audit outcome*). *Total fines*, which gives the sum of all fines and penalties applied to those businesses found underreporting taxable income has a maximum of under US$ 12,000,000, with a mean of just over US$ 56,000 and standard deviation of just over US$ 585,000. *Total audit outcome* gives the sum of *audit outcome* and *total fines*. Finally, *total audit outcome (%)* is calculated as the percentage of *total audit outcome* over the *potential tax base including fines* (defined as the sum of taxable income declared by the taxpayer and *total audit outcome* as to include tax fines).

Table 1 reveals that audits contribute a substantial amount of tax revenues in terms of uncovered taxable income underreported which amounts to 67.36 percent of the potential tax base audited in 2015 (71.47 percent including fines). Tables 2 and 3 present the same information organized by audit type grouping together the audits that are narrow in scope (desk and issue audits). In terms of the portion of underreported revenues uncovered by type of audits, comprehensive audits detect the major share of noncompliance by uncovering 92 percent of all underreporting detected while narrow-scope audits detect 8 percent.[27] In relative terms, and with respect to the corresponding self-assessment, comprehensive audits detect 60 percent of the potential tax base not including fines (62 percent including fines in the potential tax base). On the other hand, narrow-scope audits detect 72 percent of the potential tax base without fines (77 percent including fines).

[27] It is of course not surprising that comprehensive audits uncover significantly more underreported income than the narrow-scope ones. It is the impact of those audits on future compliance that is one of the concerns of this paper.

Table 1: Audits in 2015: Descriptive statistics

Variable	Obs	Measurement Unit	Mean	Std.Dev	Min	Max
Audit outcome	435	1000 US $	101.15	969.81	0	19,369.84
Audit outcome (%)	435	% Potential tax base	67.36	41.42	0	100
Total fines	435	1000 US $	56.36	585.85	0	11,621.90
Total audit outcome	435	1000 US $	157.50	1555.13	0	30,991.74
Total audit outcome (%)	435	% Potential tax base (including fines)	71.47	39.45	0	100

Note: Authors' calculations based on data provided by RRA.

Table 2: Audits in 2015: Descriptive statistics - Comprehensive audits

Variable	Obs	Measurement Unit	Mean	Std.Dev	Min	Max
Audit outcome	161	1000 US $	251.5	1584.231	0	19369.84
Audit outcome (%)	161	% Potential tax base	59.068	39.827	0	100
Total fines	161	1000 US $	143.463	957.765	0	11621.9
Total audit outcome	161	1000 US $	394.963	2541.211	0	30991.74
Total audit outcome (%)	161	% Potential tax base (including fines)	62.177	38.746	0	100

Note: Authors' calculations based on data provided by RRA.

Table 3: Audits in 2015: Descriptive statistics - Narrow-scope audits

Variable	Obs	Measurement Unit	Mean	Std.Dev	Min	Max
Audit outcome	274	1000 US $	12.799	55.78	0	854.036
Audit outcome (%)	274	% Potential tax base	72.239	41.625	0	100
Total fines	274	1000 US $	5.174	30.149	0	462.796
Total audit outcome	274	1000 US $	17.973	84.54	0	1316.832
Total audit outcome (%)	274	% Potential tax base (including fines)	76.924	38.896	0	100

Note: Authors' calculations based on data provided by RRA.

To summarize, in order to assess the impact of tax audits on CIT filers' reporting behaviour the analysis relies on different data sets matched at the (anonymised) taxpayer identification number. These are:

- The world of anonymised CIT declarations for the tax periods from 2013 to 2018.
- The world of anonymised records of completed audits undertaken by both the Large

Taxpayers Office and the Small and Medium Taxpayers Office in RRA in 2015. During the audit wave of 2015, 37.01 percent of the 435 tax audits were comprehensive, 44.6 percent were desk and 18.39 percent were issue-oriented. Notice that any taxpayer audited between $01/04/t$ and $31/03/t+1$ is classified as 'treated' in wave $t-1$ in the sense that the tax return of year $t-1$ is the last tax return reported before receiving the treatment (audit) and the tax return of year t is the first one reported after the treatment has started (that is, the first year of impact is year t). With the exception of 3 audits, which are conducted on businesses under the flat tax system, all other audits are conducted on businesses under the linear tax system (which entails both the CIT-real and CIT-lump sum tax regimes).

- Detailed information on the criteria used in audit selection, which includes the risk rules and the corresponding weighting schemes employed to assign risk scores to all tax declarations. The risk criteria utilise information across all tax bases including VAT.

The next section presents the methodology employed to estimate the causal effect of audits on the future reporting behaviour of CIT filers and discusses in detail the identification strategy.

5 Estimation strategy

To estimate the impact of audits on audited firms' future reporting behaviour—the average treatment on treated, ATT—we combine matching methods[28] with a DID approach, and we use two alternative outcome variables: The taxable income and tax liability reported as defined in Section 4 and expressed in levels. The main estimator for the aggregate[29] ATT is the Coarsened Exact Matching (CEM). CEM provides also a data preprocessing step that stratifies the sample based on relevant covariates associated with audit selection before estimating audit-type specific ATT utilizing the Inverse Probability of Treatment Weighting with regression adjustment (IPTW). Indeed, CEM possesses a set of powerful statistical properties and when used as a data preprocessing step can improve other matching methods by providing common support and improving the quality of the inferences

[28]For a discussion on the various matching methods see, among others, Stuart (2010), King et al. (2011), Imbens and Rubin (2015), and Guo and Fraser (2015).

[29]Propensity Score Matching (PSM) and Mahalanobis distance metric matching (MHD) estimators are then also employed on the CEM-stratified sample as robustness (see Appendix B.1.1). Appendix B.1.2 provides further robustness to the CEM-DID approach by estimating the aggregate ATT of audits on the whole balanced dataset—that is, without stratifying the sample through CEM—based on Synthetic DID (SDID, see Arkhangelsky et al., 2021).

drawn from these estimators.[30] The CEM procedure is intuitive. First, CEM temporarily coarsens each relevant pre-treatment variable into meaningful groups through a threshold assigned by the user based on intuitive substantive information, where it is possible, or through alternative standard binning algorithms. Subsequently, units with the same 'bin signature' (that is, with the same values) for all the coarsened variables are placed in a single stratum. And, finally, the control units within each stratum are weighted to equal the number of treated units in that stratum. Strata without at least one treated and one control unit are pruned from the data set. Each treated unit is weighted with 1 while the weights for each control unit equals the number of treated units in its stratum divided by the number of control units in the same stratum, normalized so that the sum of the weights equals the total matched sample size. By employing these weights we analyse the unpruned units through a DID approach to finally estimate the ATT through weighted fixed effects models.

Leveraging on the detailed information about the audit selection rules provided by the RRA, we have matched treated and untreated taxpayers based on their estimated risk profiles. This matching process considers both the aggregate likelihood of noncompliance and the specific sources of noncompliance for each taxpayer. Specifically, our set of pre-treatment matching variables for CEM stratification includes both the synthetic Risk Index and all its specific components that is, the risk criteria used by the RRA to flag potential noncompliance evaluated in the year of audit (2015). These are categorical variables, each flagging a specific source of potential noncompliance, involving the identification of discrepancies encountered while comparing information across tax records, tax bases and different fiscal years, as well as other behaviours that may be indicative of noncompliance. Therefore, by employing exact matching across all audit selection criteria, in addition to the synthetic risk score, our approach comprehensively addresses the taxpayers' risk profiles for noncompliance at both aggregate and granular levels, capturing nuances related to specific behaviours flagged by the RRA as indicative of different sources of potential noncompliance. In addition to these variables associated with audit selection, we also exact match on a categorical variable identifying nil-filers in 2015.[31] We

[30] Particularly, CEM has been shown to perform better than commonly used matching methods (like PSM and MHD) in reducing the initial imbalance across treatment cohorts because it targets multivariate rather than univariate imbalance reduction. CEM also reduces model dependence, estimation error, bias, variance, mean square error, and other criteria while seeking a trade-off between sample size and balance (see Iacus et al., 2011, 2012; Blackwell et al., 2009; King et al., 2011; King and Nielsen, 2019 for more details and formal proofs, and Iacus et al., 2019 for a discussion on the inference theory).

[31] A word of clarification is in order here. Audit selection is based on the product of two likelihoods, the likelihood of a business underreporting its income and conditional on underreporting (and found noncompliant) the likelihood that the audit generates some expected revenue yield. The synthetic Risk Index used relates to the former likelihood while the impact-on-revenues likelihood that is not observed relates non-linearly to the taxable income. Figure A.6 in Appendix A plots the probability of being

choose these pre-treatment covariates because they provide us with detailed information about the taxpayers' noncompliance riskiness. As a result, we obtain a good matching outcome in terms of the sought trade-off between reducing imbalance and maximizing the matched sample size.

Stratification, ideally, should involve—apart from some measures of noncompliance riskiness—exact matching on time-invariant characteristics like firms' industry and business activity, but in our case there are substantial limitations in the data regarding these dimensions due to missing data. Thus, using these variables in the stratification would reduce drastically the matched sample size below any reasonable threshold even when used in combination with fewer variables (for example, the Risk Index). Moreover, there is compelling evidence that the reliability of the information reported in these variables is doubtful. Indeed, when filling their tax returns, firms in Rwanda provide information on their business activity and industry under multiple fields and there is evidence that they often report different and completely unrelated categories across different fields even during the same fiscal year. Making use of this information in the estimation would seriously blur the identification of the actual firm's business activity and industry.[32] Generally, by adding any additional variable to the stratification procedure, in particular variables on which is necessary an exact matching or for which the share of missing values is not negligible, reduces significantly the size of the matched sample. Nevertheless, we also employ broader sets of stratification variables with respect to our baseline as robustness and obtain results that are consistent with the results of the main analysis (see Appendix B.3).

Before employing the CEM procedure to select the matched sample, restrictions are applied to the data in order the effect of one single audit to be unambiguously estimated. More specifically, a small number of outliers with effective tax rates higher than one is excluded from the control group (there are 9 observations of those) and the dataset is balanced in order to ensure that both the control and treatment units can be followed along the whole period. This ensures the pre-treatment parallel condition across cohorts can be properly checked on the selected sample. This reduces the sample to 8,710 units.

audited across deciles of the Risk Index showing it is increasing, with underlying correlation of 0.9494 between the decile of the risk score a business belongs to and the probability of being audited in that decile. Figure A.5 shows a U-shaped relationship between the probability of being audited and taxable income reported and Figure A.7 further elaborates on this by estimating the probability of being audited, by the combination of deciles of risk scores and taxable income reported (for the year before treatment is applied).

[32]There are businesses, for example, declaring to be in the manufacturing of footwear and simultaneously in the business consulting activity; or in the sewage and refuse disposal and in the retail sale of hardware; or again in the sector of manufacture of dairy products and belonging to the tour operators industry. These are all categories that do not even belong to the same broader industry.

Moreover, taxpayers who have been audited previously and/or subsequently in other waves are also excluded since not doing this would make impossible to disentangle the impact of the 2015 audit wave from the impact of other audit waves. There are 84 of those in the treatment group and 382 in the control group. With this further restriction we are left with 8,244 taxpayers to which we apply the CEM stratification procedure to select the final matched sample of 4,502 taxpayers. Table 4 summarises the selection steps.

Table 4: Description of sample selection

Sample Selection

Step	Description	Control Sample	% Δ	Audit Sample	% Δ	Total Sample	% Δ
0	Universe of CIT filers in 2015	28,739	-	435	-	29,174	-
1	Drop outliers with effective tax rate >1	28,730	99.97%	435	100.00%	29,165	99.97%
2	Balancing the panel	8,362	29.11%	348	80.00%	8,710	29.86%
3	Violation of (pre&post 2015) non-audit restrictions	7,980	95.43%	264	75.86%	8,244	94.65%
4	Final matched sample after CEM	4,287	53.72%	215	81.44%	4,502	54.61%

Note: Authors' calculations based on data provided by RRA.

CEM assesses both the reduction in the multivariate imbalance and in the univariate imbalance of pre-treatment variables through L_1 statistics introduced by Iacus et al. (2011), and reported in Table 5. Specifically, the comprehensive measure of global imbalance is based on the L_1 difference between the multidimensional histogram of pre-treatment covariates across treatment cohorts. The measures of univariate imbalance are defined analogously employing the unidimensional histograms of pre-treatment covariates (see Iacus et al. (2011) for a formal definition). In short, L_1 is bounded between 0 and 1—with higher values indicating higher imbalance—and it is an index that should be evaluated in relative rather than absolute terms by comparing the values before and after the stratification process.

Table 5 shows the performance of the CEM procedure in reducing imbalance in our sample. The overall multivariate imbalance across pre-treatment covariates after CEM almost disappears passing from 0.76 to virtually zero ($2.14e{-}15$), while maintaining a considerably high share of treated taxpayers in the final matched sample (81.44 percent, see Table 4). The reduction in pre-treatment univariate imbalance is also really pronounced in particular for the Risk Index and its components, indicating that homogeneity across treatment cohorts increases significantly in these covariates as a result of the CEM process.[33] This is visually confirmed in Figure 4 that plots the distribution of pre-treatment Risk Index before and after the CEM procedure.

Table 5: Imbalance pre and post CEM matching

Panel A: Overall imbalance, Multivariate L_1

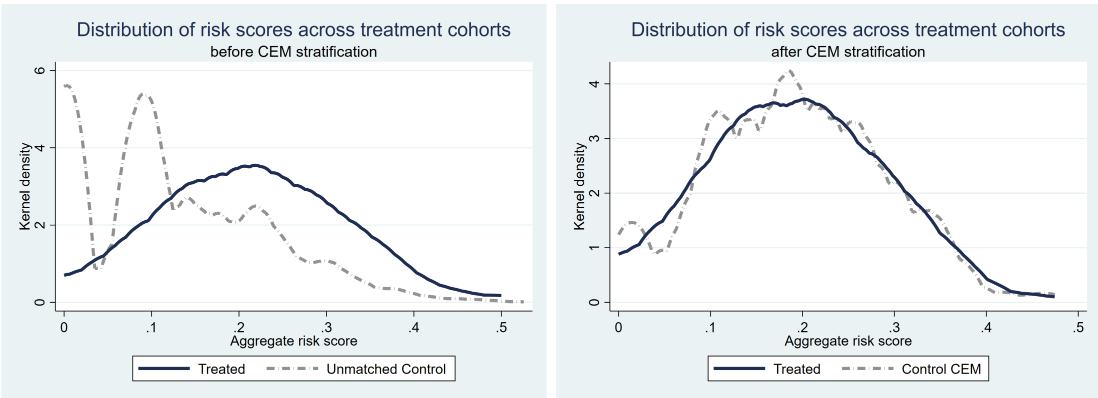
L_1 statistic pre CEM:	0.76
L_1 statistic post CEM:	2.14e-15

Panel B: Univariate imbalance

	L_1 pre CEM	L_1 post CEM
Risk Index	0.38	2.72e-15
Nil-filer	.19	8.74e-16
Risk Rules (average)	0.16	1.72e-15

Note: The table depicts L_1 statistics for multivariate and univariate imbalance as defined in Iacus et al. (2011).

Figure 4: Univariate imbalance reduction (CEM)



Note: Authors' calculations based on data provided by RRA.

Thus far we have discussed the methodology used for the estimation of the aggregate ATT for audited taxpayers independently of the type of treatment received. In order to provide an estimate of the ATT by type of audit, we aggregate audits in two main audit categories, comprehensive and narrow-scope (including desk-based and issue-oriented audits) and we perform an IPTW estimation (see, for example, Cattaneo, 2010; Wooldridge, 2002, 2007) on the CEM-matched sample.[34] IPTW estimates a type-specific

[33]Here we report the average L_1 across all the risk rules because we cannot disclose how many criteria are involved in the risk flagging procedure performed by RRA.

[34]As additional robustness checks to this analysis we also perform estimations of the ATT based on separate CEM-stratification by type of audits, also differentiating desk from issue audits within the narrow-scope category. The results obtained validate our main analysis and are discussed in Appendix B.2.

counterfactual term for any different type of audit estimating a specific ATT for any treatment. Specifically, IPTW is based on PSM and employs a multinomial logit model to estimate the propensity scores associated with a comprehensive audit (p_i^c), a desk-based audit (p_i^d) and no audit (p_i^{na}), respectively. Type-specific estimates for the counterfactual outcomes are then computed with a regression model as a weighted average of the outcomes observed for the unaudited taxpayers in the sample using as weights the ratio of the relevant treatment-specific propensity score (p_i^c or p_i^d) and the propensity score for no treatment (p_i^{na}).[35]

There is still substantial debate in the literature about how to provide valid inference when matching estimators are employed to estimate the ATT (for an insightful discussion see, for example, Iacus et al., 2019; Bodory et al., 2020). In particular, Iacus et al. (2019) argue that when ex-ante stratification solutions are employed (as, for example, for CEM) these concerns are misplaced and unaltered regression standard errors are correct. In the context of IPTW, Wooldridge (2007, 2002) has shown that ignoring the first-stage estimation of the selection probabilities when performing inference yields to more conservative standard errors than those adjusted. Moreover, the use of regression adjustment provides a double-robust technique, in the sense that implements both the estimation of the probability of treatment and the outcome regression model at once so that there is no need to correct the standard errors in the second step to reflect the uncertainty surrounding the predicted outcomes (see Cattaneo, 2010; Wooldridge, 2002, 2007). Given these premises, we provide inference by reporting robust standard errors (clustered by tax centre) for both the CEM and IPTW estimators.

The next section presents the results of the empirical analysis.

6 Results

This section presents the results, starting with the aggregate ATT, followed by the audit-type-specific aggregate ATT, and concludes with a back-of-the-envelope calculation of the return on investment (ROI) of audits. In Appendix B we present further sensitivity analysis which validates the results presented in the main text and the methodology used.[36]

[35] The set of covariates employed in the outcome model includes dummies identifying nil-filer and late reporters at the time of the audit and an indicator of the quartile of the VAT output-to-input ratio. The covariates used to estimate p_i^c, p_i^d and p_i^{na} include the risk rules specific to any type of audit, dummies for nil-filers and taxpayers reporting income from different sources, and dummies for late reporters in the last 3 years before the audit process.

[36] In particular, as already mentioned, Appendix B.1 presents the results of the analysis based on alternative matching methods (Subsection B.1.1) and SDID (Subsection B.1.2); Appendix B.2 reports the results of the analysis obtained by matching separately by type of treatment; and Appendix B.3 reports

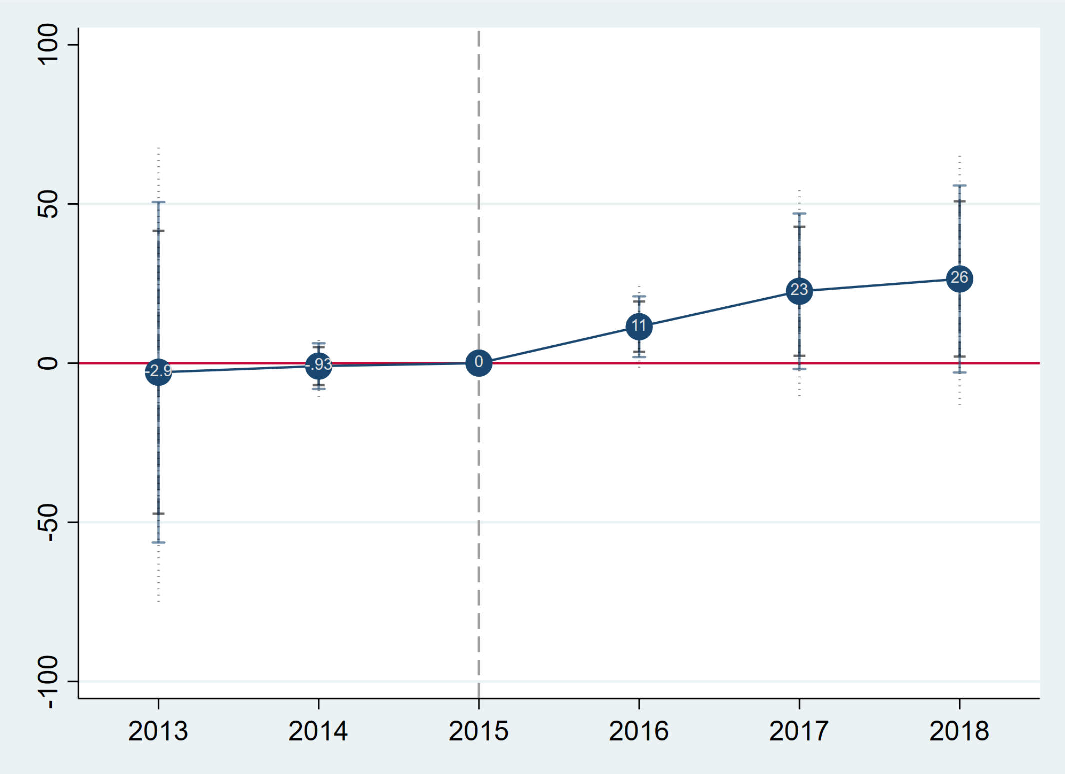
6.1 Aggregate *ATT*

A crucial assumption behind any DID analysis is the existence of a common previous trend in the outcome variable at the time of the treatment (Meyer, 1995) which, in the present context, means that one should observe a similar pattern (trend) in reporting behaviour of audited and unaudited taxpayers before treatment. To test for this, as well as estimate the period-specific audit effects on the outcome variables, we rely on Weighted Fixed Effect regressions based on panel data from 2013 through 2018 and weights obtained from our CEM stratification. In the regressions we also include both individual and year fixed effects. The excluded category is the last year before the treatment is applied (2015). Figures 5 and 6 present the results of this analysis for taxable income and tax liability, respectively, largely confirming that there is no statistically significant difference in trends between audited and matched controls before audits take place. After treatment, the estimates indicate a positive and increasing effect of audits on subsequent tax reporting behaviour of audited taxpayers for the three years after the treatment and the two outcome variables.

Quantitatively, audited taxpayers exhibit an approximate increase of US$ 11,000 in reported taxable income compared to their matched counterparts in the initial year following the audit. This difference widens to US$ 23,000 and US$ 28,000 in the second and third years after the commencement of the audit process, respectively (Figure 5). In terms of CIT liability reported in the three years after the audits, this converts to an impact of around US$ 3,600, US$ 6,100, and US$ 7,500 respectively (Figure 6). This means that CIT audits in Rwanda raise on average a total of US$ 17,246.25 by changing the behaviour of audited taxpayers over the three years after the start of the enforcement process. This amount corresponds to roughly 57 percent of the direct impact of audits in terms of detected underreporting—or approximately 20 percent of the overall detection impact of audits, inclusive of underreporting and fines—and 16.6 percent of the overall revenue collected as a result of audits (through detection and deterrence). While the deterrence impact is substantial when measured in absolute terms, its relative effectiveness compared to detection is somewhat more limited in comparison to findings from similar studies framed in the high-income context, notably Advani et al. (2021) and DeBacker et al. (2018a) which estimate the deterrence effects to be around 150 percent of the

the results of the estimation of our models obtained by applying more inclusive sets of covariates for the CEM stratification. Moreover, Appendix B.4 provides further validation of our estimation strategy by testing the sensitivity of our main results to random subsampling; and Appendix B.5 replicates the analysis on the intensive margin just for the sample of audited taxpayers with positive reportings (and their matched controls), and estimates the extensive-margin effect of audits for audited nil-filers (compared to their matched controls) through linear probability models for the likelihood to report positive outcomes after the audit process, conditional on being nil-filers at the time of the audit.

Figure 5: The Dynamic impact of audits on Taxable income reported (in levels - thousands of USD)



Note: CEM-Weighted Fixed Effect estimates of the period-specific audit effects on taxable income based on panel data 2013–2018. Individual and year fixed effects are controlled for. The excluded category is the last period before treatment (2015); 90, 95 and 99 percent confidence intervals are shown and based on robust standard errors (clustered by tax centre). Matching variables are discussed in section 5.

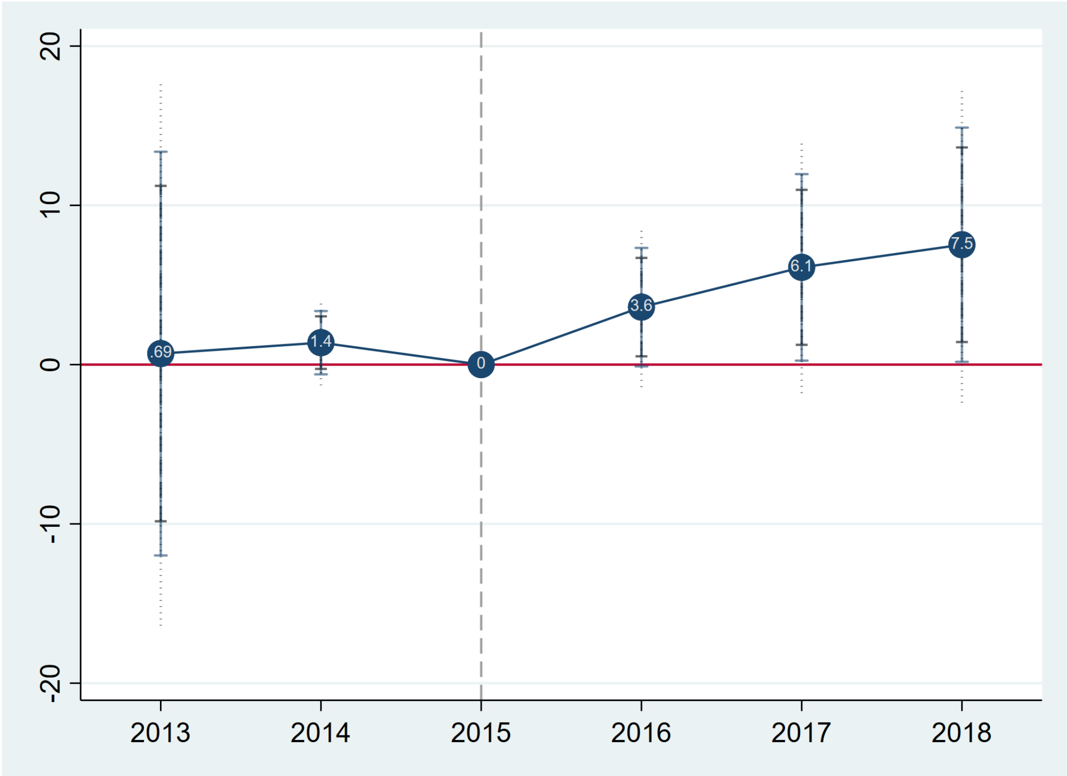
detection effect (or around 60-65 percent of total audit impact) for the UK and the US, respectively.

6.2 Comprehensive *versus* narrow-scope audits ATT

As already noted, comprehensive and narrow-scope audits are considerably different types of tax enforcement examinations.[37] Table 6 presents the results of the estimation of the ATT by audit type conducted via IPTW, showing that comprehensive audits drive the aggregate results, having a sizable pro-deterrence effect along the post-audit period. More precisely, following comprehensive audits lead to an average increase of about US$ 51,051, US$ 34,035, and US$ 129.857 in terms of taxable income reported by audited taxpayers along the three years after receiving this type of audit when compared to the control group. In terms of CIT liability reported, this impact translates to an increase of US$ 15,377,

[37] And, as alluded to in Section 3, they are generally perceived differently by taxpayers since they tend to involve different degree of deepening in the examination of declared tax items, and thus they are likely to have a different impact in deterring future noncompliance depending upon the accuracy of information conveyed to taxpayers regarding the true probability of auditing.

Figure 6: The Dynamic impact of audits on Tax liability reported (in levels - thousands of USD)



Note: CEM-Weighted Fixed Effect estimates of the period-specific audit effects on tax liability based on panel data 2013–2018. Individual and year fixed effects are controlled for. The excluded category is the last period before treatment (2015); 90, 95 and 99 percent confidence intervals are shown and based on robust standard errors (clustered by tax centre). Matching variables are discussed in section 5.

Table 6: Main Results – ATT by audit type

Dep. Variable	Taxable income reported			CIT liability reported		
Years after audit	I	II	III	I	II	III
Type of Audit	(1)	(2)	(3)	(4)	(5)	(6)
Comprehensive	51.051***	34.035***	129.857***	15.377***	10.105***	36.394***
	(5.896)	(10.908)	(15.731)	(1.808)	(3.292)	(4.333)
Narrow-scope	4.638***	-6.832***	-5.107	1.327***	-1.391***	-0.766
	(1.426)	(1.874)	(3.992)	(0.423)	(0.341)	(0.842)

Note: Robust standard errors (clustered by tax centre) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

US$ 10,105, and US$ 36,394 over the three years following the audit, with an aggregate deterrence effect totalling US$ 61,875.61. This represents 82 percent of the detection effect of these audits, when considering solely the amount of taxable income underreported, or 28 percent of the detection effect when fines are also taken into account. Expressed this in terms of total revenues collected, it corresponds to, on average, a 20 percent of total revenues collected through this type of audit.

Interestingly, narrow-scope audits have a positive impact on compliance one year after the audit, after which there is a significant counter-deterrence effect that is sufficiently strong so the aggregate impact of those audits across the three year period is negative. To be more precise, narrow-scope audits tend to have a positive impact of about US$ 4,638 (US$ 1,327) on taxable income (CIT liability) the first year after the audit and a significant counter deterrent impact of US$ 6,832 (US$ 1,391) in terms of taxable income (CIT liability) reported by businesses that experienced this type of audit. Over the third year, the point estimates (US$ -5,107 for taxable income and US$ -766 CIT liability) indicate that the negative effect tends to persist on average even though there is not enough precision to perfectly identify it.[38]

The overall revenue impact of narrow-scope audits, in relation to their detection effect—taking into consideration only the significant estimates—is -16.7 percent when considering only underreporting and approximately -0.71 percent when fines are also included in the measurement. Overall, the net average counter-deterrence effect of narrow-scope audits represent -0.72 percent of the total revenues impact of narrow-scope audits. These results are corroborated qualitatively, quantitatively and in terms of timing of the effect by a robustness analysis employing separate CEM stratification and matching for any type of audit, comprehensive, narrow-scope, and within this category desk-based and issue-oriented audits (see Appendix B.2). For narrow-scope audits, the net expected return to noncompliance is positive, even though a business has been detected to be noncompliant.

What drives the incentive to reduce compliance, having been detected as noncompliant, is certainly a complex issue. One way to rationalize this timing relates to the frequency with which businesses perceive they can be audited again and the role of information conveyed by their more mechanical nature:[39] More frequent audits make businesses more cautious in underreporting, but their mechanical nature makes them more willing to do so, given that, following the audit, they are not audited again.[40]

6.3 Returns to auditing

Both types of audits, comprehensive and narrow-scope, verify income underreported (see Tables 1-3). But comprehensive audits are compliance conducive whereas narrow-scope tax audits have, on the aggregate, a negative impact on compliance, following the tax audit. Comprehensive tax audits are also less frequent audits, verify more income under-

[38] These results are consistent with the experimental evidence provided in Kasper and Alm (2022).

[39] Though anecdotal evidence supports this, with the current data it is not a hypothesis that can be tested. But it is certainly an issue that deserves more attention.

[40] This is somewhat consistent with the restrictions applied in selecting the sample of Table 4.

reporting than narrow-scope ones, but they cost more in terms of resources allocated to them. How do these different tax audits then perform in terms of cost effectiveness? This is the question the discussion now briefly turns to.

Table 7 presents a back-of-the-envelope calculation of benefits and costs to auditing. Benefits include both the amount of underreported tax liability detected, including penalties, and the amount of tax liability associated with future compliance which corresponds to the estimated type-specific ATT over the three years after the audit. The calculation assumes that when the impact is statistically insignificant, there is no impact and so no benefit; in doing so the calculation is conservative in terms of the counter-deterrence effect of narrow-scope audits the third year after the process. The costs of audits are estimated using the salary of full time equivalent staff employed in auditing and other inspection-related activities, using data reported by the RRA to the ISORA database for the year 2019, the first available year for which this data exists (CIAT et al., 2019). The average duration of comprehensive audits is just below seven months while the average duration of narrow-scope audits is slightly more than two and a half months.

Taking now the ratio between the net benefit of each type of audits and their respective costs, the net yield to a US$ invested is around 2.5 US$ on the aggregate, but this return to investment is due entirely to comprehensive audits. Indeed, while comprehensive audits provide a yield of around 4.9 US$ per dollar invested, narrow-scope audits cost 52 cents per US$ invested.

Table 7: Return On Investment (ROI) of audits: A back-of-the envelope calculation

		Aggregate	Comprehensive	Narrow-Scope
Benefits				
	Detection	86,705.00	218,913.00	9,013.70
	Dynamic deterrence	17,246.25	61,875.61	-64.09
	Total	103,951.25	280,788.61	8,949.61
Costs		29,519.17	47,875.65	18,733.05
Net Benefit		74,432.09	232,912.96	-9,783.45
Return to Investment		2.52	4.86	-0.52

Note: Author's calculations based on data provided by RRA and estimated ATT. All values are per-audit and expressed in US$.

7 Concluding remarks

Improving tax enforcement is undoubtedly a major challenge for Revenue Authorities across the world, and is particular so for developing countries where the need for tax revenues (the only form of sustainable public finance) is significant: Just to achieve its 2030 Sustainable Development Goals in health, education, water and sanitation, roads, and electricity, Rwanda needs to spend an additional 18.7 percent of GDP per year, Gaspar et al. (2019).

By using tax administrative data for incorporated businesses and tax audits performed, this paper has investigated the role of tax enforcement in Rwanda in achieving compliance. The evidence suggests that there is in general a sizable pro-deterrence effect of tax audits on taxable income and CIT liability reported by audited businesses leading to an average increase of tax revenues of about US$17,246 over the three years after the audit process. This corresponds to about 16.6 percent of the total revenue raised by audits also including the uncovered underreporting and penalties applied, and represents approximately 5.9 percent of total CIT liability declared by all incorporated businesses in the three years after the audit process.

The results have shown that the type of audit matters. Comprehensive audits drive the pro-deterrence results while narrow-scope audits (desk-based and issue-oriented) tend to have the opposite effect starting from the second year following the audit. Interestingly, this result is consistent with the evidence provided by the recent contribution of Erard et al. (2019) for the U.S. suggesting that correspondence audits appear to be substantially less consistent in terms of improving future taxpayers' reporting behaviour and they are not, therefore, a perfect substitute for face-to-face (comprehensive) examinations. This is reflected also in the back-of-the envelope calculation regarding the return of investing a dollar on either of the two types of audits. In more general terms, and from a policy perspective, what this points to is that Revenue Authorities should pay close attention to the evaluation of their tax audits portfolio.

The analysis also suggests avenues for future research. It has abstracted, for example, from assessing the impact of past audits on compliance, following the current audit. Disentangling the impact of the current audit, from the impact of past audits is of course by no means an easy issue to explore, but it is certainly one that deserves particular attention.[41] Related to this is the role of communication and information in improving the performance of audits. Intuition, and existing evidence from the tax compliance literature (see, for example, Kirchler, 2007 and Antinuan and Asatryan, 2020), suggests that appropriate messaging could improve compliance. Whether this is the case in the context of corporate income taxes, and the distinction between comprehensive versus narrow-scope audits, it remains to be seen.

We hope to have shown that the results obtained are instructive and the issues identified merit further investigation.

[41] Along the lines of Henning et al. (2023).

Acknowledgements: We are grateful to the Management of the Rwanda Revenue Authority for supporting this research and to the many officials who have provided feedback through extensive discussions, and in particular Adrian Bizumugabe, Naphtal Hakizimana, Denis Mukama, Innocente Murasi, King Geoffrey Ngabonzima for useful comments and advice on an earlier version of this paper and Georgios Kotsogiannis for research assistance. We also thank Kieran Byrne, Christos Cabolis, Petros Dellaportas, Elizabeth Gavin, Brian Erard, Julian Jamison, Anders Jensen, Matthias Kasper, Climent Quintana-Domeque, Konstantinos Serfes, two anonymous referees and the editor (Johannes Van Biesebroeck), as well as the participants at the TARC Workshop on Tax Audits Evaluation, the TARC 8th Annual Conference, the IEB 4th Summer Workshop on the Economics of Taxation, the 19th CRETE Conference, the 77th Annual Congress of the IIPF, 7th Shadow Economy Conference, the 114th Annual NTA Conference on Taxation, the TARC-IMF 2021 Conference, the RCEA 2022 Conference and seminar participants at the Interuniversity Research Centre on Local and Regional Finance (CIFREL), Barcelona Institute of economics (IEB) and Autonomous University of Barcelona (UAB) for their helpful comments and suggestions. Kotsogiannis and Salvadori gratefully acknowledge research funding from the Economic and Social Research Council (ESRC) (Grants ES/S00713X/1 and ES/X003973/1). Salvadori also gratefully acknowledges support from the Serra Húnter Programme, the Generalitat de Catalunya [2021SGR00571] and the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S) and the Knowledge Generation Projects Programme (PID2022-137707NB-I00). The views expressed in this paper do not necessarily reflect those of the Rwanda Revenue Authority and its Management. For the purpose of open access, the authors have applied a 'Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission. The usual caveat applies.

Online Appendices: Additional figures and sensitivity analysis

Appendix A further discusses the data shedding more light on the institutional framework the analysis is based on. More importantly, Appendix B presents the results from several additional sensitivity analyses performed testing the robustness of the findings as already described in Section 5.

Appendix A: Additional figures

Table A.1 shows that the number of CIT taxpayers who submitted a tax return ('filers') has been increasing since 2013, with almost doubling in 2014 from 13,778 to 24,405 and steadily increasing thereafter by almost 10 percent year on year.

Table A.1: Number of CIT filers by fiscal year (2013-2018)

Tax period	Total number of CIT declarations
2013	13,778
2014	24,405
2015	29,174
2016	32,572
2017	36,793
2018	40,490

Note: Authors' calculations based on data provided by RRA.

This increase in the number of CIT filers is attributed to a number of factors. First, RRA introduced the possibility for small and micro businesses to opt for the above-mentioned simplified tax regimes which have progressively incentivized these types of businesses to incorporate and so broaden the CIT base. In addition to this, in 2012 there was a revision of the responsibilities of Registration and Block Management division, allowing it to focus on following up on potential unregistered taxpayers. In early 2014, there was also the establishment of the Corporate Risk Management and Modernization department and through this the introduction of a more targeted approach on audits, which served as deterrence for noncompliance. In addition, 2016 saw the establishment of the Compliance Monitoring division in Domestic Taxes Department which gave priority to the follow-up on non-filers and non-payers on regular basis.[42]

Table A.2 reports summary statistics for the main variables employed in the analysis for 2015. All monetary variables are expressed in thousands of US$, Audit is an indicator variable for audited firms, while the Risk Index is a continuous variable bounded between 0 and 1 and reflects the estimated aggregate likelihood of noncompliance synthesized by weighting all selection criteria (see also Section 5). All the Risk Rules are categorical variables representing the audit selection criteria that capture nuances related to specific

[42] Most businesses make use of electronic billing machines. A description of e-tax ecosystem is provided in Kotsogiannis et al. (2023).

behaviours flagged by the RRA as indicative of different sources of potential noncompliance.[43] Nil-filer is a dummy variable, and CIT tax regime is a categorical variable identifying the tax regime the taxpayer belongs to (see also Section 4).

Figure A.1 presents the distribution of businesses based on their reported income (expressed in natural logarithms) for selective tax periods. Given the magnitude of the nil-filers, both the complete distribution and the distribution of businesses reporting a positive income is reported. The distributions are predominantly moderately right-skewed and so the median reported income is less than the mean across all years reported.

Table A.2: Summary statistics - main variables (2015)

Variable	Obs	Mean	Std. Dev.	Min	Max
Audit	29,174	.015	0.121	0	1
Risk Index	29,174	.08	0.093	0	.526
Taxable Income	29,174	11.333	299.390	0	37,448.691
CIT liability	29,174	2.483	74.500	0	7,768.645
Risk Rules (average)	29,174	0.13	0.28	0	1
CIT tax regime	29,174	2.142	0.952	1	3
Losses reported	29,174	34.656	1,517.886	0	202,882.09
Nil-filer	29,174	.552	0.497	0	1

Note: Authors' calculations based on data provided by RRA.

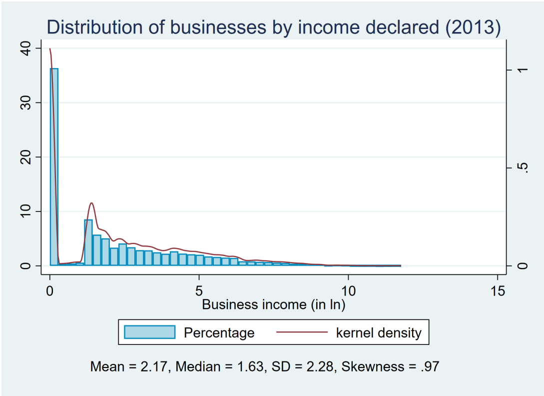
Figure A.2 reports the distribution of taxable income by deciles of population for the universe of CIT filers and by size of firms for the same selected years as in Figure A.1. Note that in terms of taxable income reported, firms in the tenth decile declare more than 90 percent of taxable income across all periods in the available data (left-hand-side panel graphs). The right-hand-side panels also show that the majority of reported income, across firm types, is reported by the top deciles of their corresponding distribution.

Figure A.3 presents the distribution of audits by type of examination and by audit wave. Across audit waves there is some variability in terms of audit types. In 2015, for example, the relative majority of audits were desk audits, whereas in 2016 the majority of tax audits is comprehensive. However, there is more stability in the relative shares of comprehensive versus narrow-scope audits (desk and issue). Indeed, both in 2014 and 2016 narrow-scope audits represent 59 percent of the total audits versus 41 percent of comprehensive and in 2015 narrow-scope audits are 63 percent of the total versus 37 percent of comprehensive. The 2013 wave is a bit more of an outlier with a 52 percent of narrow-scope audits and a 48 percent of comprehensive audits. Figure A.4 reports the distribution of audits by size of businesses together with the distribution of businesses by size across the four waves of audits. As shown in Figure A.4, most audits during 2013-2016 are performed, on average, on Small-businesses (62.87 percent of the total) following with audits on Large-businesses (19.65 percent).

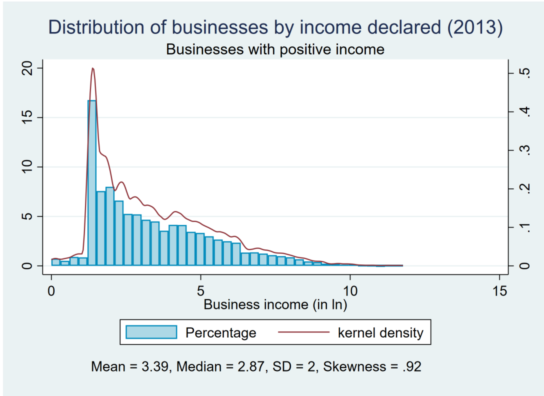
Figure A.5 shows the distribution of businesses and audits by deciles across the four audit waves (with the first six deciles having been grouped together since they include taxpayer who report nil taxable income). Audits tend to concentrate on two groups,

[43] Here we report the average figures across all the risk rules because we cannot disclose how many criteria are involved in the risk flagging procedure performed by RRA.

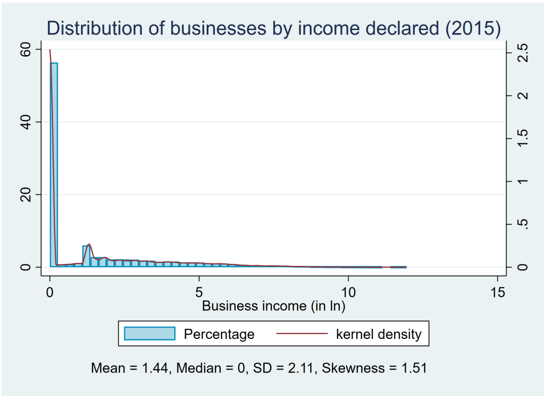
Figure A.1: Distribution of businesses by business income declared (for years 2013, 2015 and 2018)



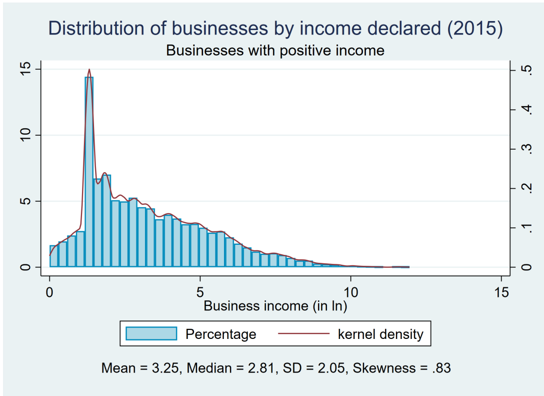
(a)



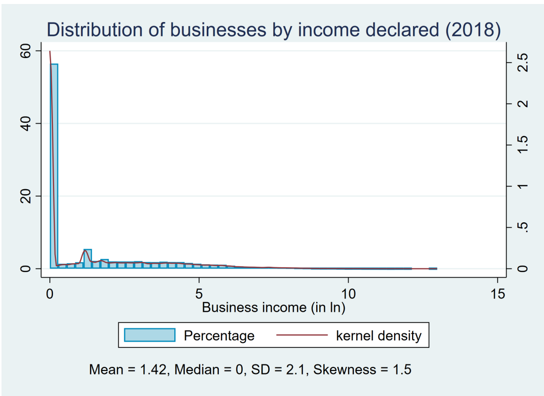
(b)



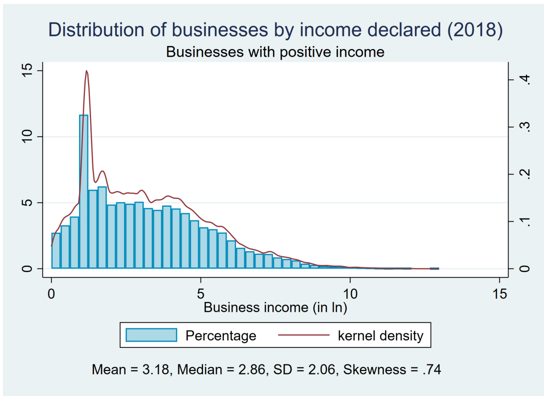
(c)



(d)



(e)

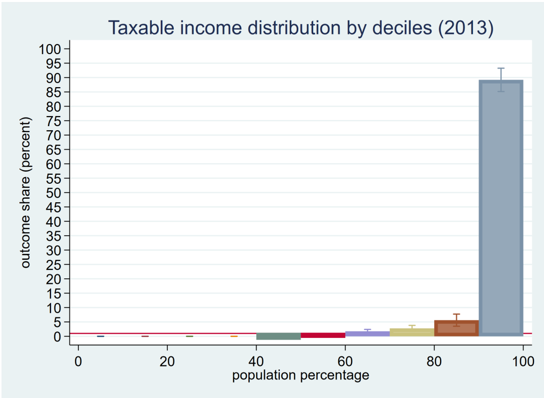


(f)

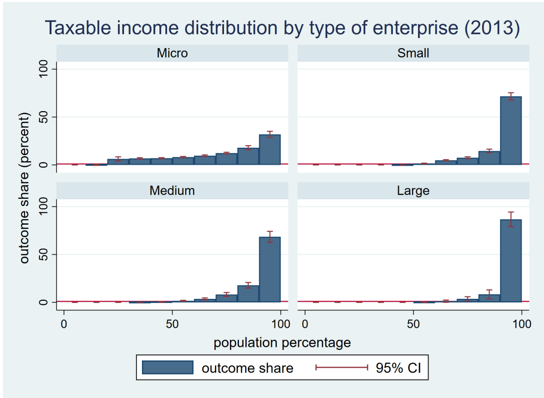
Note: Authors' calculations based on data provided by RRA.

businesses in the last decile of the taxable income and taxpayers not reporting positive tax liabilities (including nil-filers and firms reporting losses), with small businesses reporting positive taxable income generally less likely to be audited. Thus, taxable income seems to play a role in audit selection but in a non-linear fashion since the probability of being audited and the taxable income present a U-shaped relationship. Now we explore the

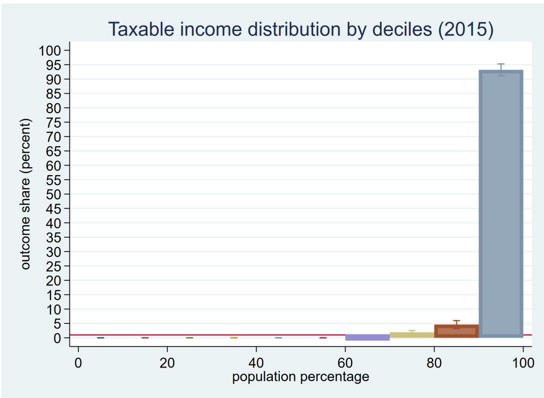
Figure A.2: Distribution of taxable income reported (for years 2013, 2015, and 2018)



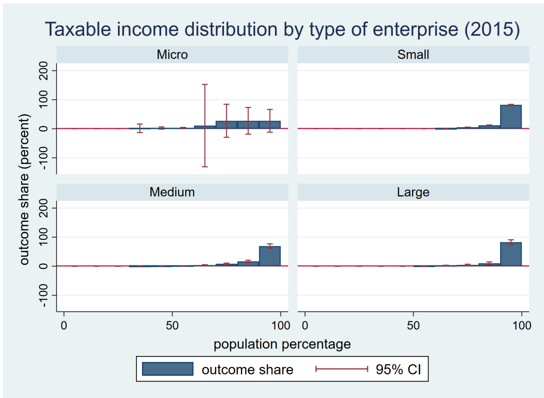
(a)



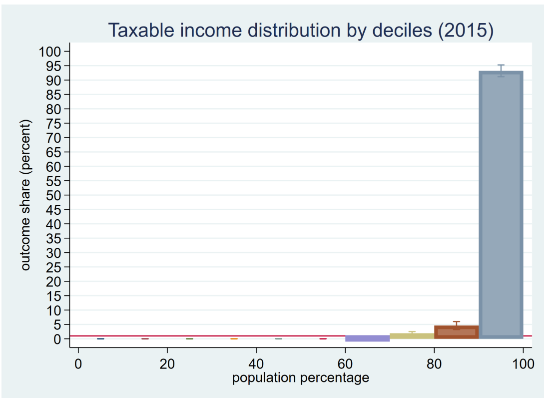
(b)



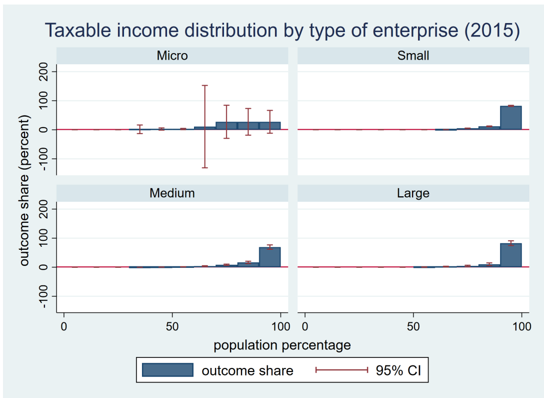
(c)



(d)



(e)



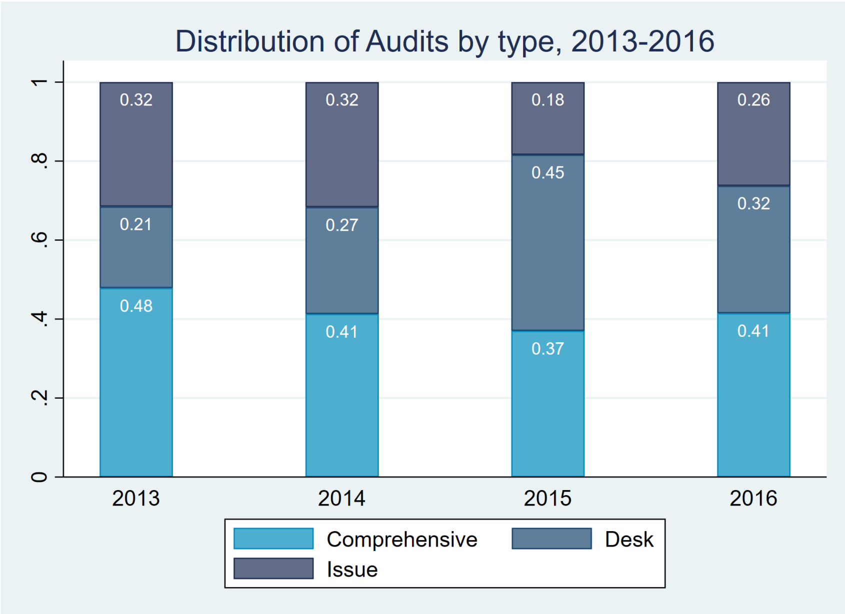
(f)

Note: Authors' calculations based on data provided by RRA.

relationship between the probability of being audited and the risk scores, the crucial dimension of our matching strategy.

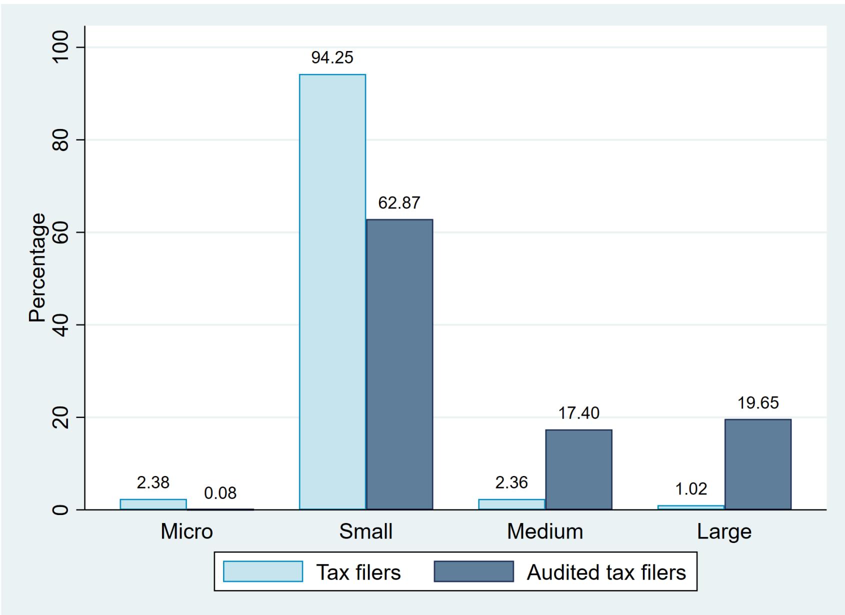
Figure A.6 plots the probability of being audited across deciles of the 2015 Risk Index which is increasing. The correlation between the deciles of the 2015 Risk Index and the probability of being audited within each decile in the 2015 audit wave is calculated to be 0.9494. What all this suggests is that the Risk Index plays an important role in

Figure A.3: Distribution of audits by type and audit wave (2013-2016)



Note: The number of audits involving incorporated business were 257 in 2013, 218 in 2014, 435 in 2015 and 337 in 2016. Authors' calculations based on data provided by RRA.

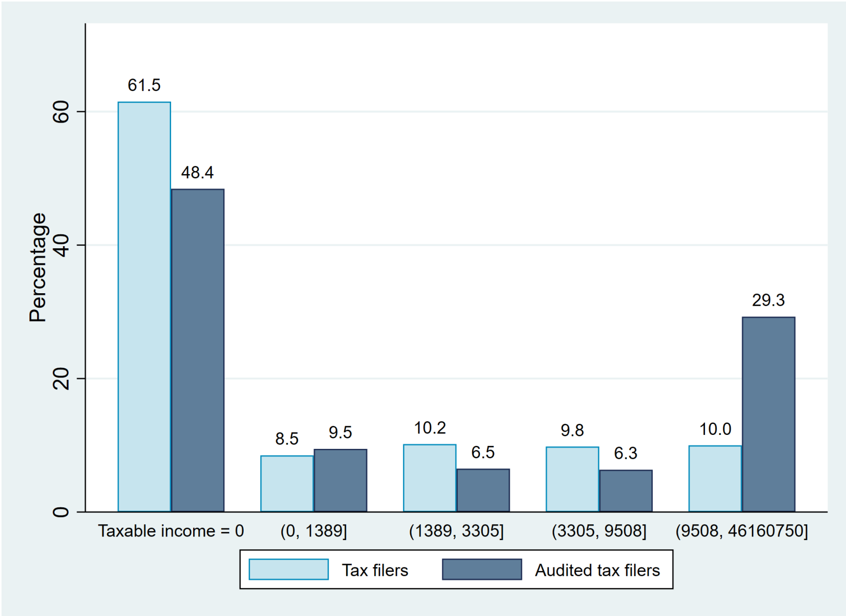
Figure A.4: Distribution of businesses and audits by size (Average 2013-2016), in percentages



Note: Authors' calculations based on data provided by RRA.

the audit selection process. As already mentioned in the main text (see footnote 31), audit selection is based on the product between the likelihood Risk Index (and so the likelihood a business to underreport its income) and (conditional on underreporting) the likelihood that the audited business generates the expected revenue yield. What this means in practice is that there exist businesses with the same likelihood Risk Index, but some are audited (and are in the treated group) whereas some are not audited (and are in the control group). This explains the lack of a clear cutoff point in the likelihood

Figure A.5: Distribution of businesses and audits by taxable income deciles (Average 2013-2016), in percentages

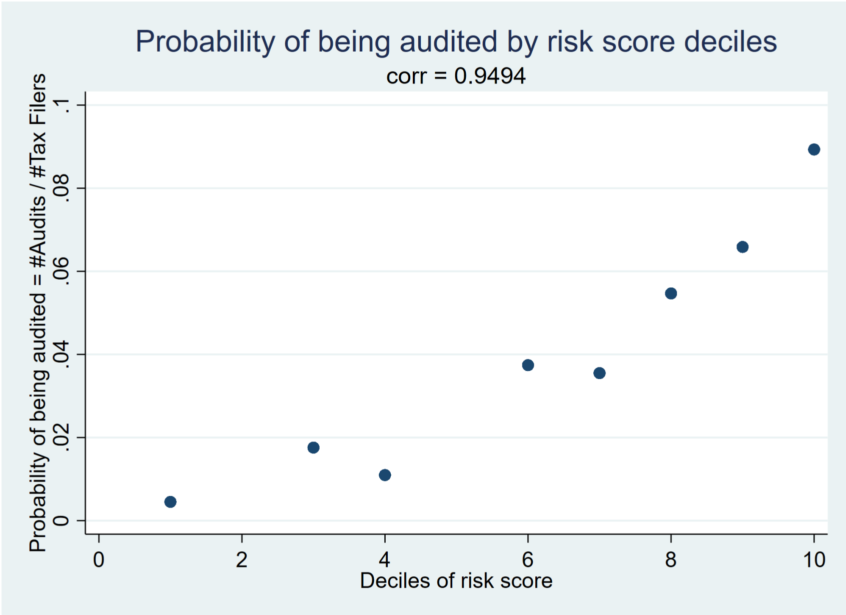


Note: Authors' calculations based on data provided by RRA.

Risk Index which determines the set of businesses to be audited.[44] Figure A.7 further elaborates on this by estimating the probability of being audited, by the combination of deciles of the likelihood Risk Index and taxable income reported. This figure suggests that the probability of being audited correlates with the combination of the risk score and taxable income in a non-linear way. Indeed, peaks of auditing probabilities can be generally found at the highest deciles of risk scores both for top income declarers (that are likely to generate yields in case of underreporting due to a scale effect) and businesses reporting nil tax liabilities (that are likely to generate yields in case they are completely concealing significant revenues) while generally a lower probability of being audited is associated with businesses declaring a relatively small amount of income regardless of the risk score decile they belong to.

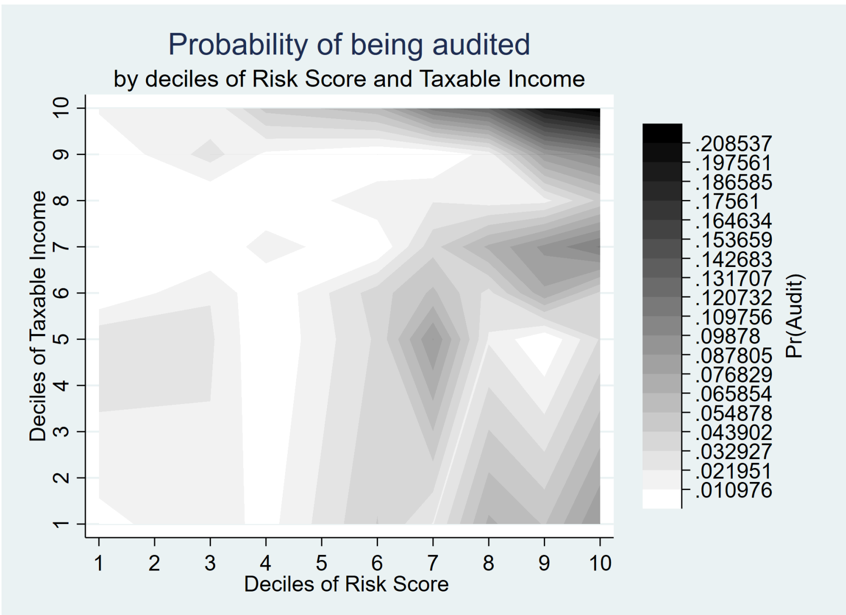
[44]For this reason a methodology that is based on a Regression Discontinuity Design cannot be employed.

Figure A.6: Probability of being audited by risk score deciles



Note: Authors' calculations based on data provided by RRA.

Figure A.7: Probability of being audited by deciles of Risk Index and taxable income reported



Note: Authors' calculations based on data provided by RRA.

Appendix B: Robustness Analysis

B.1: Alternative approaches addressing the pre-treatment parallel trend and the aggregate *ATT*

In this section of the Appendix, we test the robustness of our results on the pre-treatment parallel trend and the aggregate ATT estimated through CEM-DID to other alternative approaches. First, we employ alternative matching techniques on the same CEM-stratified sample employed for the main analysis (see Subsection B.1.1), then we adopt a less conservative approach by estimating the aggregate ATT over the three years after the audit through the Synthetic DID estimator provided by Arkhangelsky et al. (2021) (see Subsection B.1.2).

B.1.1: Alternative matching methods

As alternative specifications for the estimation of the aggregate ATT, we employ alternative matching methods on the same CEM-stratified sample and estimate weighted Fixed Effects models based on the resulting weights. Namely, we use the Kernel PSM estimator, the Kernel MHD estimator and the Nearest Neighbour MHD estimator (NN). The combination of CEM preprocessing with other matching methods has been proven to enhance them by providing common support and improving the quality of inferences (see Iacus et al., 2011 and also Section 5).[45] Based on a set of pre-treatment covariates,[46] Kernel PSM estimates the propensity score through a discrete choice model (in our case logit) where the dependent variable reflects assignment to audit treatment and subsequently estimates the counterfactual term as a simple weighted average of the outcomes with weights reflecting the similarity across participants.[47] More precisely, the Kernel PSM estimator assigns weights based on the expression $W(i,j) = \frac{G\left(\frac{p_j - p_i}{h}\right)}{\sum_{k \in I_0} G\left(\frac{p_k - p_i}{h}\right)}$, where $G(.)$ is a kernel function, h is the number of observations falling into the bandwidth, I_0 is the identification function for the control group and p_i, p_j and p_k are the estimated propensity scores for treated i unit and the j and k control units, respectively. The Kernel MHD estimator is conceptually similar to the Kernel PSM estimator but employs the Mahalanobis distance metric for pre-treatment covariates,[48] instead of the distance between propensity scores $p_j - p_i$, to define the similarity between units i and j and to calculate weights. Finally, the Nearest Neighbour MHD estimator also utilizes the Mahalanobis distance metric to

[45] Indeed, as recently discussed in the literature (see, for example, King et al., 2011, and King and Nielsen, 2019), these estimators do not generally guarantee any level of imbalance reduction and can even increase imbalance and model dependence.

[46] We employ a sequential selection process in order to select the final set of pre-treatment covariates to estimate the propensity score based on their predictive power. The final set includes the 2015 synthetic Risk Index and all its components (risk rules), a dummy for nil filers, categorical variables for the three tax regimes and ventiles of taxable income and cit liability reported by taxpayers in 2013.

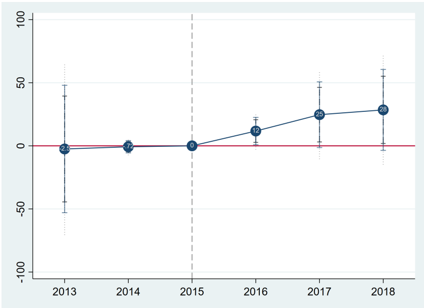
[47] More generally, several alternative PSM algorithms may be used to weight units instead of the Kernel estimator in the second step of the PSM procedure. However, as suggested by Heckman et al. (1997, 1998), matching based on local polynomial regressions, like the Kernel estimator, are more efficient because they construct weighted average counterfactuals based on all control group units. For this reason we have selected the Kernel estimator within the PSM class.

[48] The Mahalanobis distance metric for covariate X and units i and j can be defined in the following way: $d_{i,j} = \sqrt{(X_i - X_j)S^{-1}(X_i - X_j)}$, where S is the sample covariance matrix of X.

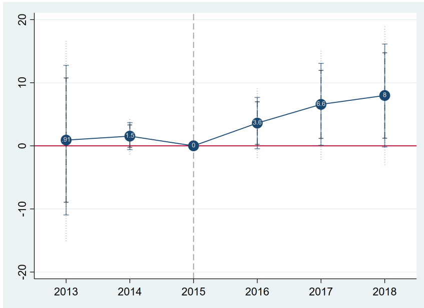
synthesize similarities across treatment cohorts but it assigns weights in a simpler way: for any treated unit i, this method assigns a weight $W(i, j) = 1$ to the control members j with the lowest levels of the Mahalanobis distance for the relevant covariates and 0 to the others. We use the twenty nearest neighbours in the control group for any treated taxpayer to build the conterfactual. For both the MHD and NN we use as matching variables the synthetic Risk Index, a dummy for nil-filers, the ventiles of taxable income reported by taxpayers in 2013 and dummies for the tax regime the taxpayers belong to.

Figure B.1 reports the results of this analysis that are qualitatively and quantitatively similar to those presented in the main text.

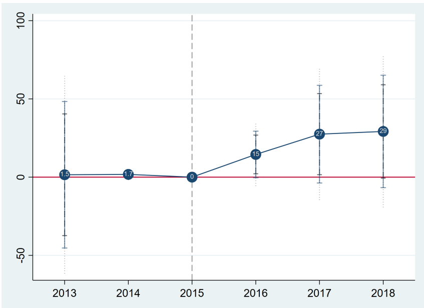
Figure B.1: The Dynamic impact of audits: alternative matching methods



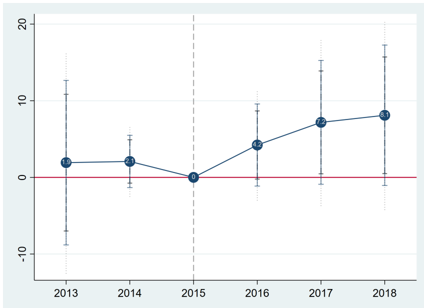
(a) PSM - Taxable income



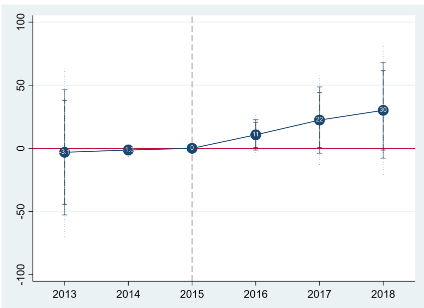
(b) PSM - Tax liability



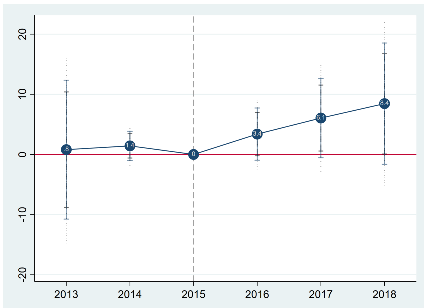
(c) MHD - Taxable income



(d) MHD - Tax liability



(e) NN - Taxable income



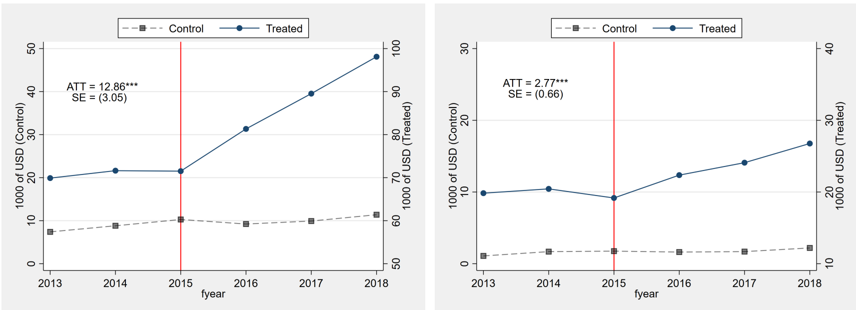
(f) NN - Tax liability

Note: Weighted Fixed Effect estimates of the period-specific audit effects on taxable income and tax liability based on panel data 2013–2018 and different matching estimators. Individual and year fixed effects are controlled for. The excluded category is the last period before treatment (2015); 90, 95 and 99 percent confidence intervals are shown and based on robust standard errors (clustered by tax centre).

B.1.2: Synthetic DID

As a further robustness to our CEM-DID approach, here we report the estimation of the aggregate ATT of audits for the whole post-treatment period based on Synthetic DID (SDID, see Arkhangelsky et al., 2021). We perform this on the whole balanced dataset composed by 264 audited taxpayers and 7,980 control businesses (as in step 3 of Table 4, that is without stratifying the sample through CEM). Results presented in Figure B.2 corroborate the existence of a pre-treatment parallel trend and lead to an overall statistically significant post-treatment effect that resembles results obtained through our main estimation strategy, although the overall effect estimated through SDID for the whole post-treatment period is lower in absolute value when compared with the same figure obtained through CEM-DID (see also Table B.1, columns 1 and 3 for taxable income, columns 4 and 6 for CIT liability).

Figure B.2: The Dynamic impact of audits: Synthetic DID



(a) SDID - Taxable income (b) SDID - Tax liability

Note: This graphs depict the trends in reporting behaviour for both outcome variables over time for audited taxpayers and the relevant weighted average of control taxpayers using the whole balanced dataset. Estimates of the aggregate audit ATT for the whole post-treatment period (2016-2018) based on SDID along with estimated standard errors are also reported. We use the "placebo method" standard error estimator (with 200 repetitions) discussed in Section IV of Arkhangelsky et al. (2021); * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.1: Aggregate post-treatment ATT (2016-2018), SDID and CEM comparison

Dep. Variable	Taxable income reported			CIT liability reported		
Estimator	SDID	SDID-CEM	CEM	SDID	SDID-CEM	CEM
	(1)	(2)	(3)	(4)	(5)	(6)
ATT (2016-2018)	12.864***	20.360***	20.161*	2.771***	4.363***	5.749**
	(3.047)	(2.851)	(10.244)	(0.656)	(0.723)	(2.733)
Observations	49464	27000	27012	49464	27000	27012

Note: Standard errors are reported in parentheses, in particular "placebo method" standard error estimator (with 200 repetitions) discussed in Section IV of Arkhangelsky et al. (2021) is used for SDID while Robust standard errors (clustered by tax centre) are reported for CEM; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

This difference is attributable to the stricter selection of the control group we applied in our main analysis (step 4 of Table 4) by exact matching on all the covariates relevant for audit selection. Instead, what SDID does is reweighting the whole set of unexposed control units (not just those selected through CEM) to make their time trend parallel (but not necessarily identical) to audited taxpayers' pre-treatment reportings, then apply a DID analysis to this reweighted panel. This is further clarified by Table B.1 which compares the aggregate 2016-2018 post-treatment impact of audits alternatively estimated through SDID on the whole balanced dataset (columns 1 and 4, as in Figure B.2), by using SDID on the CEM-stratified sample (columns 2 and 5), and by using CEM-Weighted Fixed Effect model (columns 3 and 6). The estimates of CEM-stratified SDID and CEM-DID are much more similar also from a quantitative point of view.

B.2: Separate CEM stratification by type of treatment

In order to estimate the type-specific ATT we rely on IPTW. This method creates type-specific counterfactuals by estimating separate probability of being comprehensively audited, narrowly audited and not audited and using the ratio between the probability of being audited with any type of audit divided by the probability of not being audited as weight in the estimation of the type-specific ATT (see Section 5). Here, we present a robustness to this methodology by producing separate CEM-stratification for any audit category. This entails performing the matching process twice—one for the comprehensive audits and one for the narrow-scope audits—in order to estimate the treatment effects of the two types of audits separately.

We also do the same exercise by differentiating within the narrow-scope category and estimate separately the ATT for desk-based and issue-oriented audits. This exercise provides a test to our hypothesis that the two types of audits can be clustered under the same category of narrow-scope audits. A priory, we grouped desk audits and issue-oriented audits together because they are homogeneous in two aspects on which we focus on. First, they are both narrow in their scope given that they tend to focus on a single aspect, a single tax base and a single tax period. Second, they are similar in their low intensity and duration. Both these aspects are crucial in shaping how these audits are perceived differently from comprehensive audits by the taxpayers, given the information they convey.

The results of this exercise, reported in Table B.2, are consistent to those obtained estimating the ATT specific to any type of audit by employing the IPTW on the whole sample also validating our hypothesis that desk and issue audits lead to the same negative impact, with the same timing. The estimated impact is also very similar in terms of magnitude of the effect and more precise on the third year after the audit.

Table B.2: Main Results – ATT by audit type (Separate CEM stratification by type)

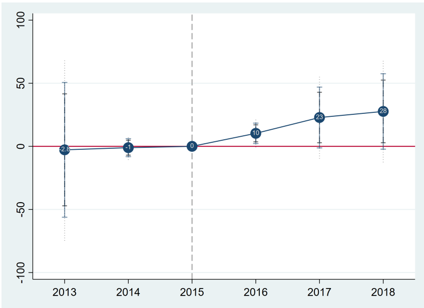
Dependent Variable	Taxable income reported			CIT liability reported		
Years after the audit	I	II	III	I	II	III
Matching estimator	(1)	(2)	(3)	(4)	(5)	(6)
Comprehensive	24.339*	60.731**	81.961**	8.007*	15.221***	22.023**
	(14.176)	(24.841)	(41.317)	(4.641)	(5.879)	(10.767)
Narrow-scope	5.507***	-9.935***	-7.414*	1.538***	-2.340***	-1.516
	(1.550)	(2.167)	(4.457)	(0.401)	(0.391)	(1.008)
Desk	4.080*	-8.831**	-6.779	1.116***	-1.856**	-1.020
	(2.461)	(4.057)	(5.768)	(0.429)	(0.867)	(1.281)
Issue	6.544***	-10.104**	-6.600***	1.878***	-2.560*	-2.092***
	(2.038)	(4.190)	(1.834)	(0.562)	(1.388)	(0.548)

Note: Robust standard errors (clustered by tax centre) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

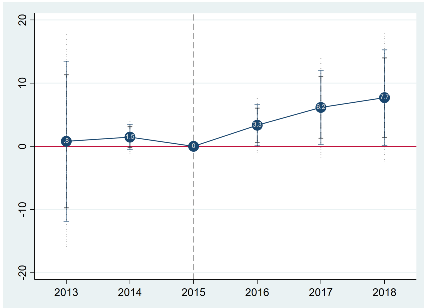
B.3: Alternative sets of CEM covariates

As a further robustness check we select our matched sample by applying the CEM stratification to less parsimonious sets of covariates and subsequently implement our baseline models. We have employed several more inclusive sets of covariates for CEM selection as alternative to the one presented in Section 5. Here we present the results of three of them. Figure B.3 present the results on the aggregate while Tables B.3-B.5 report the results of the analysis by type of audit.

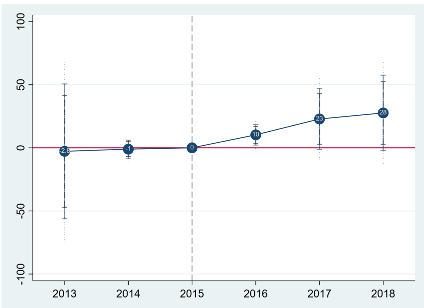
Figure B.3: The Dynamic impact of audits: alternative sets of CEM covariates



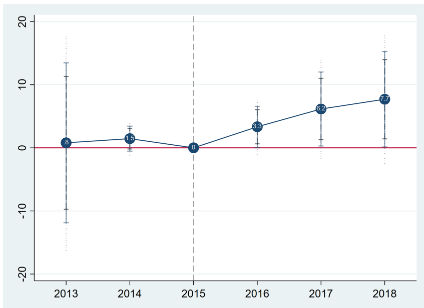
(a) Taxable income - Set II



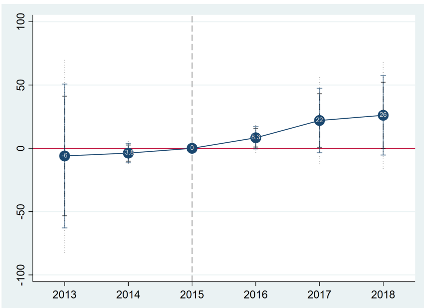
(b) Tax liability - Set II



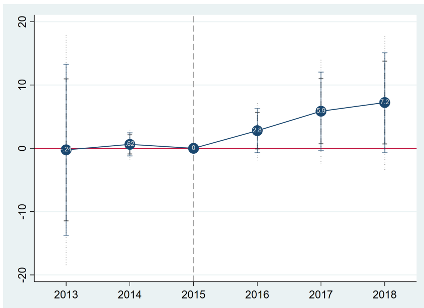
(c) Taxable income - Set III



(d) Tax liability - Set III



(e) Taxable income - Set IV



(f) Tax liability - Set IV

Note: CEM-Weighted Fixed Effect estimates of the period-specific audit effects on taxable income and tax liability based on panel data 2013–2018 and different set of matching variables (as described in Appendix B.3). Individual and year fixed effects are controlled for. The excluded category is the last period before treatment (2015); 90, 95 and 99 percent confidence intervals are shown and based on robust standard errors (clustered by tax centre).

Set II of matching covariates includes the initial set of control variables described in Section 5 and adds a categorical variable identifying the three types of tax regimes the taxpayer belongs to (as described in Section 4). Exact matching on this variable indirectly accounts for the size and the sophistication of the business.

The matched set of observations includes 214 treated units (81 percent of the pre-treatment units) and 3,450 untreated units (43.2 percent), slightly less than the sample used for the main analysis but still a comparable sample size. After CEM, the multivariate imbalance measure on the whole broader set of matching variables reduces to less than half of the already negligible residual imbalance obtained after matching on the initial set of matching variables. This is a substantial reduction, considering that it is inherently more difficult to reduce imbalance simultaneously on an extended set of variables. Also the univariate imbalance metrics reduce substantially. By using a larger set of stratification covariates that includes Set II of covariates and adds the amount of losses reported by the business, we are still able to match 214 treated units (81 percent of the pre-treatment units) and 3,449 untreated units (43.2 percent). With Set VI of marching variables, we also add dummy variables for nil-filers in 2013 and 2014 being able to match 203 treated units (76.9 percent) and 2,728 control units (34.2 percent).

Table B.3: Main Results – ATT by audit type (using Set II of matching covariates)

Dep. Variable	Taxable income reported			CIT liability reported		
Years after audit	I	II	III	I	II	III
Type of Audit	(1)	(2)	(3)	(4)	(5)	(6)
Comprehensive	51.415***	35.744***	132.549***	15.502***	10.512***	36.838***
	(7.021)	(10.383)	(15.589)	(2.155)	(3.179)	(4.265)
Narrow-scope	3.928**	-5.643***	-2.826	1.154***	-1.109***	-0.544
	(1.872)	(1.803)	(3.140)	(0.438)	(0.340)	(0.838)

Note: The different sets of matching variables are described in Appendix B.3. Robust standard errors (clustered by tax centre) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.4: Main Results – ATT by audit type (using Set III of matching covariates)

Dep. Variable	Taxable income reported			CIT liability reported		
Years after audit	I	II	III	I	II	III
Type of Audit	(1)	(2)	(3)	(4)	(5)	(6)
Comprehensive	51.391***	35.728***	132.497***	15.494***	10.507***	36.824***
	(7.037)	(10.416)	(15.579)	(2.160)	(3.189)	(4.264)
Narrow-scope	3.924**	-5.631***	-2.812	1.153***	-1.105***	-0.540
	(1.874)	(1.793)	(3.123)	(0.437)	(0.337)	(0.833)

Note: The different sets of matching variables are described in Appendix B.3. Robust standard errors (clustered by tax centre) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The results of the analysis corroborate our main findings both qualitatively and quantitatively. In particular, across the different stratification solutions there is little variation in terms of the absolute value of the estimated effect, its significance and its trend in the three years following the audit. Also regarding the type-specific ATT the analysis confirms the results both in terms of the sign of the estimated effects and their

Table B.5: Main Results – ATT by audit type (using Set IV of matching covariates)

Dep. Variable	Taxable income reported			CIT liability reported		
Years after audit	I	II	III	I	II	III
Type of Audit	(1)	(2)	(3)	(4)	(5)	(6)
Comprehensive	63.948***	41.713***	164.632***	19.260***	12.381***	46.085***
	(10.005)	(13.590)	(31.346)	(2.993)	(4.128)	(8.739)
Narrow-scope	3.470	-5.519**	-3.668	1.019**	-1.031**	-0.787
	(2.193)	(2.326)	(4.507)	(0.435)	(0.460)	(1.234)

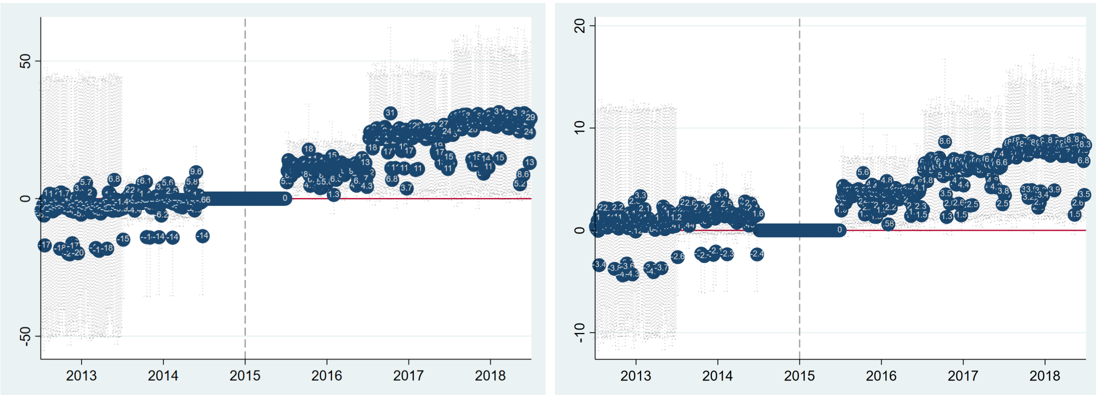
Note: The different sets of matching variables are described in Appendix B.3. Robust standard errors (clustered by tax centre) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

magnitude (see Tables B.3-B.5). This whole robustness analysis shows that performing the CEM stratification on a larger set of covariates reduces (marginally) the sample size but do not alter our results which are confirmed both qualitatively and quantitatively, always showing the same sign and generally very similar magnitudes in the estimated effects of audits both on the aggregate and by type of audits.

B.4: Robustness on outliers

In order to test how sensitive are our results to outliers, we draw 100 subsamples out of our matched dataset in a way that each subsample keeps a different randomly selected 95 percent of the treated sample and all their matched controls, and we perform the estimation of the ATT on those samples. Figure B.4 plots the results of these subsampling analysis for the aggregate ATT. The pre-treatment coefficients are never statistically significant, confirming the existence of a pre-treatment parallel trend, while the post-treatment estimates are generally statistically significant and resemble both qualitatively and quantitatively those estimated for the whole sample (Figures 5 and 6). Figure B.5 presents the results of the same exercise for the type-specific impact of audits. The estimates obtained corroborate those presented in Section 6.2 both qualitatively and quantitatively. The first year the impact of both narrow-scope and comprehensive audits is positive and significant and tend to concentrate close to the effect estimated for the whole sample. The second and third year after the enforcement started, the impact of comprehensive audits remains positive and significant while narrow-scope audits present the same impact both in terms of its negative sign and magnitude to the one reported in the main analysis presenting the same level of significance as in Table 6. Overall, this random subsampling exercise seems to corroborate that our main results are not sensitive to outliers.

Figure B.4: Random subsampling on 95 percent of treated (and their matches) - Aggregate results

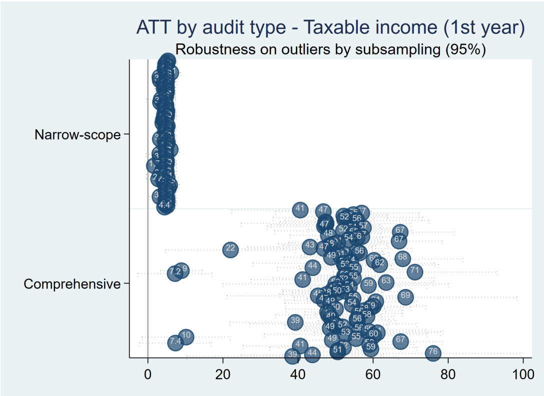


(a) Taxable income

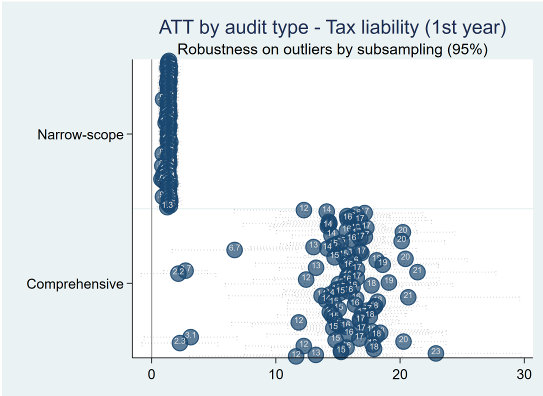
(b) Tax liability

Note: These graphs represent the results of the estimation of the aggregate ATT obtained using CEM in 100 random subsamples including 95 percent of the treated taxpayers and their matched control units. On each subsample, CEM-Weighted Fixed Effect models are performed to obtain estimates of the period-specific audit effects on taxable income and tax liability. Individual and year fixed effects are controlled for. The excluded category is the last period before treatment (2015); 90 percent confidence intervals are shown and based on robust standard errors (clustered by tax centre).

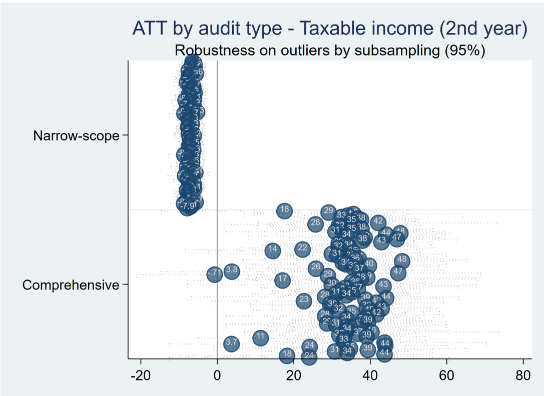
Figure B.5: Random subsampling on 95 percent of treated (and their matches) - by type of audit



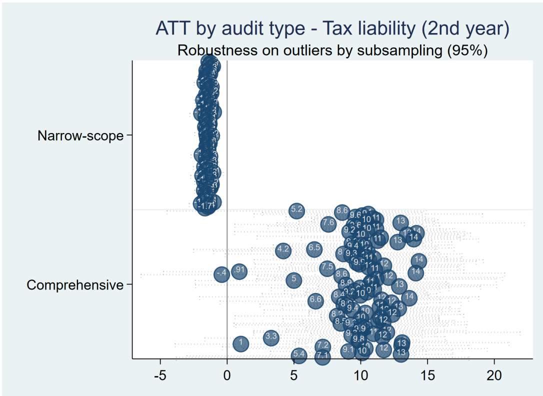
(a)



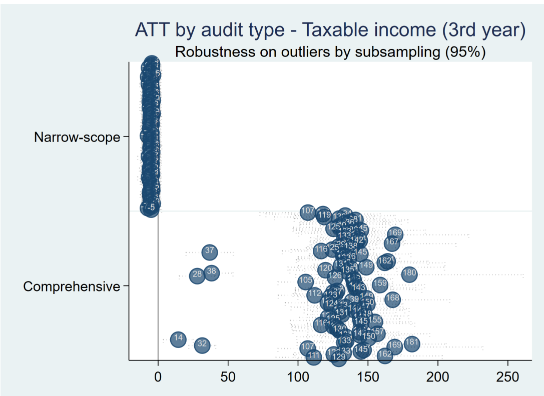
(b)



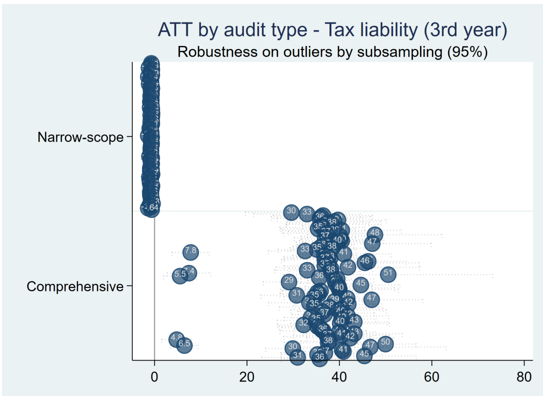
(c)



(d)



(e)



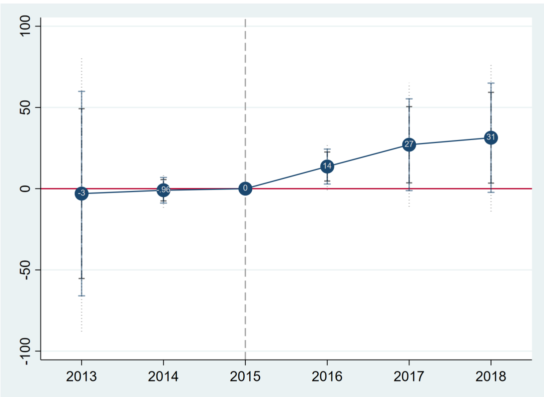
(f)

Note: These graphs represent, by outcome variable (taxable income and CIT liability) and year of the impact (I, II, and III), the results of the estimation of the type-specific ATT obtained using IPTW in 100 random subsamples including 95 percent of the treated taxpayers and their matched control units. Estimates are depicted with 90 percent confidence intervals.

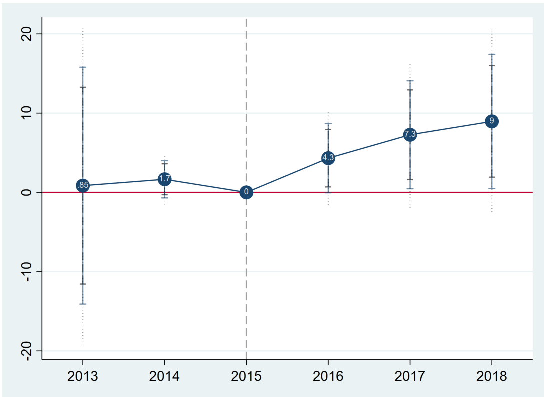
B.5: Audited tax filers vs. Nil-filers

Our matched sample is composed by 215 audited taxpayers, of which 181 are reporting positive taxable income and liabilities while 34 are Nil-filers. Our set of matching variables involves a dummy variable identifying nil-filers, assuring that across the treatment cohorts taxpayers are matched within the same reporting category. Nevertheless, in order to shed more light about the specific impact of treatment on the behaviour of these different types of taxpayers, here we replicate the analysis presented in the main text on the intensive margin (Figures 5-6 and Table 6) just for the sample of audited taxpayers with positive reportings (and their matched controls), and we perform an estimate of the extensive-margin effect of audits for audited nil-filers (compared to their matched controls). This entails estimating linear probability models for the likelihood to report positive outcomes after the audit process, conditional on being nil-filers at the time of the audit.

Figure B.6: The Dynamic impact of audits: Intensive margin for audited Taxpayers reporting positive taxable income and CIT liabilities



(a)



(b)

Note: CEM-Weighted Fixed Effect estimates of the period-specific audit effects on taxable income and CIT liabilities based on panel data 2013–2018 and restricted to taxpayers reporting positive outcomes at the time of the audit. Individual and year fixed effects are controlled for. The excluded category is the last period before treatment (2015); 90, 95 and 99 percent confidence intervals are shown and based on robust standard errors (clustered by tax centre). Matching variables are discussed in section 5.

Figure B.6 and Table B.6 present the results on the intensive margin for audited taxpayers with positive reportings, which are qualitatively and quantitatively coherent with those obtained for the whole matched sample, with similar level of significance. The analysis of the extensive margin for nil-filers reveals an aggregate positive impact of audits on the likelihood that a nil-filer will report positive outcome following the audit process (Table B.7).[49]

[49] Given the limited number of treated taxpayers in this cluster, we do not have enough power to differentiate the extensive-margin impact by type of audit.

Table B.6: Main Results – ATT by audit type (intensive margin for positive tax filers)

Dep. Variable	Taxable income reported			CIT liability reported		
Years after audit	I	II	III	I	II	III
Type of Audit	(1)	(2)	(3)	(4)	(5)	(6)
Comprehensive	43.977***	24.557	126.485***	13.230***	7.424	34.852***
	(5.282)	(23.959)	(25.685)	(1.597)	(7.186)	(7.020)
Narrow-scope	5.213**	-8.391***	-6.413	1.619***	-1.716***	-0.920
	(2.042)	(2.687)	(4.569)	(0.559)	(0.578)	(1.034)

Note: Robust standard errors (clustered by tax centre) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Matching variables are discussed in section 5.

Table B.7: Aggregate post-treatment ATT (2016-2018): Extensive margin for Nil-filers

Dep. Variable	Positive taxable income reported	Positive CIT liability reported
	(1)	(2)
ATT (2016-2018)	0.140***	0.141***
	(0.028)	(0.028)
Firm Fixed Effects	YES	YES
Year Fixed Effects	YES	YES

Note: CEM-Weighted Linear probability models. Dependent variables are the probability to report a positive value of Taxable income and CIT liability. Robust standard errors (clustered by tax centre) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Advani, A., Elming, W., Shaw, J., 2021. The Dynamic Effects of Tax Audits. The Review of Economics and Statistics 0, 1–45.

Allingham, M.G., Sandmo, A., 1972. Income Tax Evasion: A Theoretical Analysis. Journal of Public Economics, 1, 323–328.

Alm, J., Malézieux, A., 2021. 40 Years of Tax Evasion Games: A Meta-Analysis. Experimental Economics, 0, 1–52.

Antinuan, A., Asatryan, Z., 2020. Nudging for Tax Compliance: A Meta-Analysis. CESifo Working Paper, 8500.

Arkhangelsky, D., Athey, S., Hirshberg, D.A., Imbens, G.W., Wager, S., 2021. Synthetic Difference-in-Differences.

Atwood, T., Drake, M., Myers, J.N.and Myers, L., 2012. Home Country Tax System Characteristics and Corporate Tax Avoidance: International Evidence. The Accounting Review, 87, 1831–1860.

Balán, P., Bergeron, A., Tourek, G., Weigel, J., 2021. Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in D.R. Congo. CEPR Working Paper, 15138.

Beer, S., Kasper, M., Kirchler, E., Erard, B., 2020. Do Audits Aeter Future Noncompliance? evidence On Self-Employed Taxpayers. CesIfo Economic Studies, 66, 248—-264.

Bergeron, A., Tourek, G., Weigel, J., 2020. The State Capacity Ceiling on Tax Rates: Evidence from Randomized Tax Abatements in the DRC. Mimeo.

Best, M., Shah, J., Waseem, M., 2021. The Deterrence Value of Tax Audit: Estimates from a Randomised Audit Program. Mimeo.

Blackwell, M., Iacus, S., King, G., Porro, G., 2009. Cem: Coarsened Exact Matching in Stata. The Stata Journal, 9, 524–546.

Bodory, H., Camponovo, L., Huber, M., Lechner, M., 2020. The Finite Sample Performance of Inference Methods for Propensity Score Matching and Weighting Estimators. Journal of Business & Economic Statistics 38, 183–200.

Brockmeyer, A., Smith, S., Hernandez, M., Kettle, S., 2019. Casting a Wider Tax Net: Experimental Evidence from Costa Rica. American Economic Journal: Economic Policy, 11, 55–87.

Cattaneo, M.D., 2010. Efficient Semiparametric Estimation of Multi-Valued Treatment Effects Under Ignorability. Journal of Econometrics 155, 138–154.

CIAT, IOTA, IMF, OECD, 2019. International Survey on Revenue Administration (ISORA). URL: https://data.rafit.org.

DeBacker, J., Heim, B.T., Tran, A., Yuskavage, A., 2015. Legal Enforcement and Corporate Behavior: An Analysis of Tax Aggressiveness after an Audit. Journal of Law and Economics, 58, 291–324.

DeBacker, J., Heim, B.T., Tran, A., Yuskavage, A., 2018a. Once Bitten, Twice Shy? The Lasting Impact of Enforcement on Tax Compliance. Journal of Law and Economics, 61, 1–35.

DeBacker, J., Heim, B.T., Tran, A., Yuskavage, A., 2018b. The Effects of IRS Audits on EITC Claimants. National Tax Journal, 71, 451–484.

Eberhartinger, E., Safaei, R., Sureth-Sloane, C., Wu, Y., 2020. Are Risk-Based Tax Audit Strategies Rewarded? An Analysis of Corporate Tax Avoidance. Mimeo.

Ebrahim, A., Kisanga, E., Swema, E., Leyaro, V., Mhede, E.P., Mdee, E., Palviainen, H., Pirttilä, J., 2021. The effects of a risk-based approach to tax examinations: Evidence from Tanzania. WIDER Working Paper 2021/150.

Erard, B., Kirchler, E., Olsen, J., 2019. Audit impact study: The specific deterrence implications of increased reliance on correspondence audits, in: Taxpayer Advocate Service: Annual Report to Congress 2019, pp. 257–268.

Erikson, W., 1969. A Note on the Posterior Mean of a Population Mean. Journal of the Royal Statistical Society: Series B (Methodological) 31, 332–334.

Finley, A.R., 2019. The Impact of Large Tax Settlement Favorability on Firms' Subsequent Tax Avoidance. Review of Accounting Studies, 24, 156–187.

Gaspar, V., Amaglobeli, D., Garcia-Escribano, M., Prady, D., Soto, M., 2019. Fiscal Policy and Development: Human, Social, and Physical Investment for the SDGs. IMF Staff Discussion Note n. 19/03.

Guo, S., Fraser, M., 2015. Propensity Score Analysis: Statistical Methods and Applications. Advanced Quantitative Techniques in the Social Sciences, SAGE Publications. Second Edition.

Heckman, J.J., Ichimura, H., Todd, P., 1998. Matching as an Econometric Evaluation Estimator. Review of Economic Studies, 65, 261–294.

Heckman, J.J., Ichimura, H., Todd, P.E., 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. Review of Economic Studies, 64, 605–654.

Henning, D., Kotsogiannis, C., Salvadori, L., Pirttilä, J., 2023. Audits and tax compliance: Evidence from Uganda. Mimeo.

Hoopes, J.L., Mescall, D., Pittman, J.A., 2012. Do IRS Audits Deter Corporate Tax Avoidance? The Accounting Review, 87, 1603–1639.

Iacus, S.M., King, G., Porro, G., 2011. Multivariate Matching Methods That Are Monotonic Imbalance Bounding. Journal of the American Statistical Association, 106, 345–361.

Iacus, S.M., King, G., Porro, G., 2012. Causal Inference without Balance Checking: Coarsened Exact Matching. Political Analysis, 20, 1–24.

Iacus, S.M., King, G., Porro, G., 2019. A theory of statistical inference for matching methods in causal research. Political Analysis 27, 46–68.

Imbens, G.W., Rubin, D.B., 2015. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press, USA.

Kasper, M., Alm, J., 2022. Audits, audit effectiveness, and post-audit tax compliance. Journal of Economic Behavior & Organization 195, 87–102.

Kasper, M., Rablen, M.D., 2023. Tax compliance after an audit: Higher or lower? Journal of Economic Behavior & Organization 207, 157–171.

King, G., Nielsen, R., 2019. Why propensity scores should not be used for matching. Political Analysis, 27, 435–454.

King, G., Nielsen, R., Coberley, C., Pope, J.E., Wells, A., 2011. Comparative Effectiveness of Matching Methods for Causal Inference. Mimeo.

Kirchler, E., 2007. The Economic Psychology of Tax Behaviour. Cambridge University Press.

Kleven, H.J., Knudsen, M.B., Kreiner, C.T., Pedersen, S., Saez, E., 2011. Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. Econometrica, 79, 651–692.

Kotsogiannis, C., Salvadori, L., 2024. Tax Administration and Fiscal Capacity: Challenges in Developing Countries. Mimeo.

Kotsogiannis, C., Salvadori, L., Karangwa, J., Innocente, M., 2023. Do Tax Audits and Electronic Billing Machines Generate Tax Compliance Spillovers? Forthcoming in TARC Discussion Papers.

Kotsogiannis, C., Serfes, K., 2014. The Comparison of ad Valorem and Specific Taxation under Uncertainty. Journal of Public Economic Theory 16, 48–68.

Kotsogiannis, C., Serfes, K., 2016. Optimal Performance Reward, Tax Compliance and Enforcement. Economic Theory Bulletin 4, 325–345.

Lediga, C., Riedel, N., Strohmaier, K., 2020. Tax Enforcement Spillovers – Evidence from South Africa. Mimeo.

Li, W., Pittman, J.A., Wang, Z., 2019. The Determinants and Consequences of Tax Audits: Some Evidence from China. The Journal of the American Taxation Association, 41, 91–122.

Løyland, K., Øvrum, A., 2017. The Effect of Targeted Audits on Immigrant Workers' Tax Compliance. Working Paper.

Mascagni, G., Monkam, N., Nell, C., 2016. Unlocking the Potential of Administrative Data in Africa: Tax Compliance and Progressivity in Rwanda. ICTD Working Paper, 56.

Mascagni, G., Santoro, F., Mukama, D., Karangwa, J., Hakizimana, N., 2022. Active Ghosts: Nil-filing in Rwanda. World Development 152.

Mendoza, J.P., Wielhouwer, J.L., Kirchler, E., 2017. The Backfiring Effect of Auditing on Tax Compliance. Journal of Economic Psychology, 62, 284–294.

Meyer, B.D., 1995. Natural and Quasi-experiments in Economics. Journal of Business & Economic Statistics, 13, 151–161.

OECD, AUC, ATAF, 2021. Revenue Statistics in Africa 2021. OECD Report.

Reinganum, J.F., Wilde, L.L., 1986. Equilibrium Verification and Reporting Policies in a Model of Tax Compliance. International Economic Review 27, 739–760.

RRA, 2019. Tax Statistics in Rwanda, Fiscal Year 2017/18. Second Edition, June 2019, Kigali.

Slemrod, J., 2019. Tax Compliance and Enforcement. Journal of Economic Literature, 57, 904–954.

Stuart, E.A., 2010. Matching Methods for Causal Inference: A Review and a Look Forward. Statistical Science, 25, 1–21.

Tourek, G., Dada, R., 2021. Peer Information and Social Norms in Tax Compliance: Evidence from Rwanda. Mimeo.

Waseem, M., 2021. The role of withholding in the self-enforcement of a value-added tax: Evidence from pakistan. Review of Economics and Statistics, 0, 1–44.

Wooldridge, J.M., 2002. Inverse Probability Weighted M-estimators for Sample Selection, Attrition, and Stratification. Portuguese Economic Journal 1, 117–139.

Wooldridge, J.M., 2007. Inverse Probability Weighted Estimation for General Missing Data Problems. Journal of Econometrics 141, 1281–1301.