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

TARC DISCUSSION PAPER 035-21

CHRISTOS KOTSOGIANNIS, TARC, UNIVERSITY OF EXETER
LUCA SALVADORI, TARC, UNIVERSITY OF EXETER
JOHN KARANGWA, RWANDA REVENUE AUTHORITY
THEONILLE MUKAMANA, RWANDA REVENUE AUTHORITY

w www.tarc.exeter.ac.uk
t +44 (0) 1392 723464
TARC is an ESRC Research Centre

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

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

First version: September 7, 2021
This version: February 16, 2022

*Abstract*: Making use of a unique administrative data set for the period 2013-2018 consisting of the universe of administrative filings in Rwanda this paper investigates the impact of tax audits on incorporated businesses' reporting behaviour. Using matched-Difference-In-Difference the evidence suggests that the average aggregate effect—estimated across different matching approaches—corresponds to an increase of 20.7% in Corporate Taxable Income (CTI) reported by audited businesses the year after receiving the audit that in turn corresponds to an increase of 12.3% in Corporate Income Tax (CIT) paid by those taxpayers. The results also suggest that the type of audit matters. While comprehensive (face to face) tax audits have a significant pro-deterrence effect with an average increase of 28.5% (24.6%) in CTI reported (and CIT payable), narrow scope (desk-based) tax audits, exhibit a counter-deterrent effect on future reporting behaviour leading to a sizeable reduction of 23.5% (9.5%) in CTI (and CIT payable) reported by taxpayers that experienced this kind of tax audit. The implication of this is that narrow scope audits are not a substitute for comprehensive audits, and doing more of the former and less of the latter might have a negative impact on tax compliance.

*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, Streatham Court, Rennes Drive, EX4 4PU, England, UK.

[b]CESIfo, Munich, Germany.

[c]IEB, Barcelona, Spain.

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

[†]Corresponding author. Email: l.salvadori@exeter.ac.uk.

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, Murasi Innocente, Denis Mukama, and King Geoffrey Ngabonzima for useful comments and advice on an earlier version of this paper. We also thank Kieran Byrne, Christos Cabolis, Petros Dellaportas, Elizabeth Gavin, Brian Erard, Julian Jamison, Anders Jensen, Matthias Kasper, Climent Quintana-Domeque, Konstantinos Serfes, 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, and seminar participants at the Interuniversity Research Centre on Local and Regional Finance (CIFREL) and Barcelona Institute of economics (IEB) for their helpful comments and suggestions. Kotsogiannis and Salvadori gratefully acknowledge research funding from the Economic and Social Research Council (ESRC) [under grant ES/S00713X/1]. The views expressed in this paper do not necessarily reflect those of the Rwanda Revenue Authority and its Management. The usual caveat applies.

# 1 Introduction

Recent estimates have it that achieving the Millennium Development Goals[1] 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 successful efforts to increase tax revenues and boost domestic revenue mobilisation over the last few years in developing countries continue, revenue mobilization is now even more challenging, and pressing, as governments announce much needed relief measures to ease the impact of the COVID-19 pandemic.[2] Amongst the important key policy reforms for improving revenue mobilization is strengthening tax administration capacity, an issue which has become a key reform priority for many countries around the world during the last two decades.

An integral part of tax compliance is understanding the role of operational audits. Tax audits can affect a taxpayer's behaviour in three ways. First, there is an effect (direct and specific) through verification of tax liabilities, and any required adjustment identified, on the audited tax returns. Second, there is a deterrence effect (direct) on the future compliance behaviour of taxpayers who have been audited, and, finally, there is a deterrence effect (indirect) through those who have not experienced an audit but belong to the wider network of the audited taxpayers.[3] Recent literature on audit assessment—based on high-income countries data—has focused on the estimation of the direct and specific deterrence effects and thus on the impact of audits on future tax compliance by the audited taxpayers. The impact of tax audits on future compliance turns out to be ambiguous: Audits might contribute positively on tax compliance through taxpayers updating their perceived probability of being audited again upwards[4]—as in, among others, Kleven et al. (2011),

[1]The Sustainable Development Goals (SDGs) were adopted by the member countries of the United Nations (UN) and they are intended to guide the global development agenda through 2030 by focusing on the view that development needs to be economically, socially, and environmentally sustainable. The COVID-19 pandemic is expected to reverse the progress made on education, health and living standards for the first time since 1990. For more details see UN-DP (2020).

[2]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% of GDP. During this period external debt costs also increased by 1.1% of GDP a figure that is expected to rise considerably as a consequence to the adverse impact of the COVID-19 pandemic, OECD/AUC/ATAF (2021).

[3]This can take the form of a social network, as in Gamannossi degl'Innocenti and Rablen (2020) and Boning et al. (2018), where network centrality emerges as a focal element of the audit strategy, or the form of an economic network as in VAT where businesses are linked through a sequence transactions, an issue discussed in Pomeranz (2015). Spatial spillovers are also implicitly possible as a result of geographic proximity to the audited taxpayer an issue discussed in Lediga et al. (2020).

[4]In the standard tax compliance model, a taxpayer reports its tax liability to the tax authority trading off the benefit and cost of tax evasion (Allingham and Sandmo, 1972). The cost of evasion is that with some probability the tax authority discovers evasion and recovers the evaded amount together with a

Løyland and Øvrum (2017), DeBacker et al. (2018a,b), Beer et al. (2020), and Advani et al. (2021)—but their impact might also be negative (an effect which has been colloquially described as the 'bomb-crater' effect) if taxpayers perceive the probability of them being audited again (or 'struck-twice') low. Until recently, that tax audits can have a negative impact on future tax compliance was a theoretical possibility.[5] 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 of the 'bomb-crater' effect in the sense that fiscal controls may lead to an increase in noncompliance. By using IRS tax administrative data for self-employed taxpayers, Erard et al. (2019) find a counter-deterrent effect of 7.3% of desk audits in the years immediately after the audit and of 8.3% the year after, while they find a sizeable pro-deterrence impact of audits on compliance for comprehensive audits (of around 40% increase in reported tax liability). This distinction in audit types will be at the heart of this paper. We return to this shortly below.

Recently, research has begun to pay attention to developing countries, and there is now an emerging literature analysing tax compliance issues from a number of different perspectives (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[6] and Best et al., 2021 on the deterrence value of VAT audits in Pakistan). But the issue of evaluating tax audits is, rather surprisingly, neglected given its importance for much needed revenue mobilisation.[7] And this is the objective of this paper: to investigate

penalty for the misreported tax liability. The likelihood that this occurs is in general unknown to the taxpayers, even though, it is reasonable to assume, that they do form beliefs from any past interaction with the tax authority which they update following new information received. This is an issue that is taken up and further conceptualised in Section 3.

[5]For more discussion on this see the meta-analysis on tax compliance of Alm and Malézieux (2021) who, covering 70 tax compliance laboratory studies, show that enforcement variables (such as audits and fines) perform differently on the extensive and intensive margins while fiscal variables (for example, flat tax regimes, tax rates and tax amnesties) have an unambiguously negative impact on tax compliance. Antinuan and Asatryan (2020) in a meta-analysis covering 55 studies involving randomised control trails find that on average the effects of such interventions are modest, increasing the probability of compliance by only 1.5-2.5 percentage points.

[6]See also Mascagni et al. (2016, 2019) 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% on the first year of implementation. The emergence recently of empirical contributions using administrative data has been fostered by better access to administrative tax-return data, driven by the willingness of tax authorities to engage with academic researchers in their pursuit, following public demand for more efficient and accountable tax authorities.

[7]Perhaps the neglect comes from the fact that quality tax administrative data has been until recently

the impact of auditing on deterring future noncompliance of incorporated businesses[8] in Rwanda, paying particular attention to evaluating the impact of the different tax audits (something we return to shortly below) 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 (Slemrod, 2019).

Focusing on Rwanda presents a unique opportunity in understanding how well tax audits perform in a developing country who is embracing reforms and whose economy over the last decade has been growing steadily, earning the country a reputation as one of Africa's fastest-growing economies. Despite the particularly challenging social circumstances which Rwanda has gone through, it has implemented major reforms successfully to the extend that the Rwandan tax administration is now recognized as a benchmark for other tax administrations in Africa. During the last ten years Rwanda has seen a sizeable growth in taxation revenues: tax-to-GDP ratio was 11.8% in 2009 and reached 15.9% in 2017, RRA (2020).

Rwanda's long term developments goals are embedded in the Rwanda Vision 2050 with its main objective to transform the economy into a knowledge-based, service-oriented economy with a middle-income status by 2050. To achieve this objective over the years the Rwandan government has invested significantly in tax administration capacity and has implemented a broad set of reforms. Part of the transformation has been to strengthen tax auditing capacity, monitoring of non-filers and non-payers, and investing in IT capabilities, including the enhancement of electronic billing machines (EBM).[9]

To address the issues discussed above use is made of a unique administrative data for the period 2013-2018 which consists of administrative tax filings in Rwanda and tax audits performed by the RRA. More specifically, the analysis utilises the universe of CIT and Value Added Tax (VAT) anonymized tax declarations for the period 2013-2018, the universe of risk-based anonymized audit data for the audit wave in 2015[10] as well as detailed information on the risk rules and criteria with the corresponding risk weighting scheme employed by RRA to prioritise CIT filers for audit selection.[11] Tax audits are

difficult to come by. We return to this later on and in Section 5.

[8]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.

[9]This paper, and the collaborative research with the RRA underlying it, is part of the compliance strategy of the RRA. By understanding how well operational audits perform in terms of revenue yield and future compliance the RRA contributes to its strategic vision regarding much needed revenue mobilization.

[10]While the focus of this paper is on audit wave 2015, due to completeness and quality of the relevant data, additional analysis is performed using the audit wave of 2013. The results corroborate the main findings and are reported in Appendix E.1.

[11]Risk-based audits combine specific elements in a tax filer's return to determine the likelihood of under-reporting of corporate tax income as well as VAT obligations. The information used in risk-based

prioritized by RRA through a risk-based assessment (and computerized) procedure which implies that 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 robust *pro-deterrence* effect of audits on Corporate Taxable Income (CTI) and CIT payable reported one year after the audit process. Specifically, the estimates reveal an average increase of around 20.7% in CTI reported by audited taxpayers after receiving the audit which corresponds to an increase of roughly 12.3% in CIT payable reported by these taxpayers.[12] The effect is lower in magnitude during the following years but it is not statistically significant. Interestingly, the estimated aggregate effect seems to be driven by the change in behaviour of audited taxpayers who have been identified by RRA as noncompliant.

RRA performs three types of audits, 'desk-based', 'issue-oriented' (which are generally also desk-based and narrow in scope) and 'comprehensive'. Narrow-scope audits are used by RRA for reviewing tax filed and other documents submitted to the Revenue Authority and they tend to focus narrowly on a single item of the tax return. Desk audits are considered impersonal and they are also less costly to undertake for the Revenue Authority (as well as for businesses). Comprehensive audits, on the other hand, are in-depth, in-person and across tax bases examinations that involve the inspection of the accounting books as well as transactions of the audited firms. Narrow-scope audits cover around of 60% of tax audits, a number that is likely to increase following COVID-19 and the restrictions imposed by the Revenue Authority on face to face meetings in an effort to mitigate the spread of the virus and its consequences. Understandably, the level of (minimal) intensity associated with narrow-scope audits has meant that they are high on the policy recommendations of the international organizations providing technical assistance to low-capacity tax administrations. The nature of the tax audits, however, and their diverse degree of intensity, raises the issue of whether narrow-scope and comprehensive tax audits might be perceived differently by the taxpayer and conceivably they might have a different impact on businesses' compliance behaviour. As it will be shown the evidence points to this.

More specifically, the results show that comprehensive audits drive the pro-deterrence result with an average increase of approximately 28.5% in CTI reported corresponding to

audits stem from operational experience but also knowledge obtained through past audits. Although random audits do generate valuable information on compliance behaviour predominantly audits are risk-based as they are considered to be a more efficient audit selection strategy in terms of their opportunity cost and therefore their revenue generation than cases selected randomly (OECD, 2006).

[12] CTI reported and CIT payable reported are of course related but not in an exact way. For this reason the analysis reports estimates on both outcomes.

24.6% more CIT payable reported by audited taxpayers after receiving this specific type of audit. Interestingly, and perhaps surprisingly, desk and issue audits (narrow-scope) tend to have a non-significant effect the first year after the start of the audit process and have an opposite (counter-deterrent) effect from the second year following the audit. This translates to a reduction of about 23.5% in CTI reported by taxpayers that experienced this kind of audit. In terms of CIT payable reported these audits correspond to a reduction of 9.5%.[13] A possible explanation for this behaviour[14] is that taxpayers who experience low intensity audits—and on single issues such as VAT refund, non-deductible expenses, invoices not declared, insufficient documents to assess tax liability—and have underreported on fields not examined under a desk audit, they may revise their expected gain from noncompliance upwards (through lowering the expected probability of auditing) thereby increasing noncompliance, following the audit. To put it differently, audits on specific issues (coupled with any other information received by businesses regarding the likelihood of them being audited) reveal to businesses that the Revenue Authority does not hold accurate information regarding their true tax underreporting, which determines the audit probability. This does not suggest that narrow-scope audits are not a desirable instrument. But it does suggest that the presumption that if Revenue Authorities do more of those, and less of comprehensive, tax compliance will improve is incorrect: narrow-scope tax audits do not seem to be a perfect substitute for comprehensive examinations.

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 taxpayer's 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 the strand of the literature that evaluates the (direct) impact of tax audits on audited taxpayers. Like most of the recent contributions employing different methodologies across different contexts (see, for example, Kleven et al., 2011; Løyland and

[13]The results obtained on audit type relate, conceptually, to the recent contribution utilising US data finding 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).

[14]And consistent with taxpayers updating their beliefs regarding the quality of information on their true income the tax authority possesses or its ability to process the information effectively. In low capacity tax administrations it might be a combination of both (see Besley et al., 2013).

Øvrum, 2017; DeBacker et al., 2018a,b; Beer et al., 2020, and Advani et al., 2021), this paper finds an aggregate significant average positive impact of tax audits on the future reporting behaviour of audited taxpayers. Unlike them, however, the focus of this paper is on the corporate income tax base and developing countries and also on the compliance impact of the different types of audits. Investigating the deterrence impact of tax audits on these three margins is the main contribution of this paper.[15]

The existing, and surprisingly limited, literature on businesses' future tax reporting behaviour, following tax audits, provides mixed evidence. To the best of our knowledge there are only two contributions that directly evaluate the future impact of audits on corporate behaviour using business level administrative data.[16] On the one hand, 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. In particular, these authors show that after firms have been audited they significantly increase taxes paid, reduce their book-tax differences, and also reduce their income-decreasing discretionary accruals. On the other hand, tax audits might also increase corporate tax aggressiveness, a finding reported in DeBacker et al. (2015) who, using US 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. The reason that has been put forward to explain this finding is that large taxpayers are more informed regarding audit risks.[17] This, however, might not necessarily be the case in developing countries where the set of incorporated businesses may include a significant share of small and micro businesses whose behaviour is likely to be comparable to that of individual taxpayers.[18]

Importantly, the contribution of this paper is also in extending the somewhat scant evidence on the specific deterrence effect of audits on future corporate tax underreporting but within the broader context of a developing country. Indeed, more generally the analysis of the future (direct and indirect) deterrence effects of audits has been largely overlooked in this context.

Recently, by focusing on VAT, Best et al. (2021) exploit a national program of

[15]Slemrod and Gillitzer (2014) provide an excellent discussion on the issue. The same is true across different legal entities since, typically, their reporting requirements are different.

[16]Additional contributions based on different approaches also provide mixed evidence (see Atwood et al., 2012; Hoopes et al., 2012; Finley, 2019, and Eberhartinger et al., 2020).

[17]There is some evidence that supports this view, showing that corporations—compared to individual taxpayers—present a higher level of tax sophistication and they are willing to take the risk associated with aggressive tax planning (Armstrong et al., 2019). An implication of this is that large businesses tend to respond to different social norms and networks (Hasan et al., 2017) than small businesses and generally face different costs of noncompliance (Hanlon et al., 2007; Hanlon and Slemrod, 2009 and Jacob et al., 2021).

[18]We return to this later on in the paper.

randomized VAT audits in Pakistan to investigate how much evasion audit uncovers and how much evasion it prevents through changing behaviour of businesses. They find that although tax audits uncover a substantial amount of evasion (the evasion rate among firms in the bottom three quartiles is more than 100%) they do not deter future compliance. Based on interviews they had with tax auditors, these authors rationalise their findings 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. Within the context of developing world, in a recent contribution Lediga et al. (2020) provide some evidence on network deterrence effects of audits. By investigating whether corporate tax audits in South Africa have an impact on neighboring (that is, those that are in close proximity) businesses, the authors find that enforcement effects are short-run and level off two years after the audit has taken place. The revenue impact of audits is in the region of 6.5% of the amount verified underreported. Our results suggest that the direct revenue impact of the specific deterrence effect of corporate audits in Rwanda corresponds to 11.9% of the amount of underreporting detected for that tax base. One would probably expect that the direct impact of audits is more significant than the indirect (through the network), so these revenue outcomes seem fairly consistent.

In addition to comprehensive audits, RRA also rely on desk-based and issue-oriented audits. Interestingly, the evidence suggests that these audits have the opposite impact to the one intended for: they increase noncompliance. This is an issue which has been rather neglected in the literature. Notable exception to this is Erard et al. (2019) who show that while risk-based face-to-face audits always deter future noncompliance, correspondence audits tend to have a counter-deterrent effect, with taxpayers reducing reported taxes by 6% to 15% over the two years following the audit. Arguably, this is an issue that deserves more attention given its importance for the optimal allocation of resources across the types of tax audits undertaken by Revenue Authorities, particularly in the developing world.

As noted earlier, the existing literature finds a positive effect of audits which is short-lived: it is very strong in the first years after the audit but rapidly converges towards the pre-audit levels in following years (Kleven et al., 2011; Løyland and Øvrum, 2017; DeBacker et al., 2018a,b; Beer et al., 2020, and Advani et al., 2021).[19] Interestingly,

[19] Specifically, the impact tends to be driven by self-reported income (see, for example, Kleven et al., 2011 and DeBacker et al., 2018a). While third-party reporting and tax withholding makes it very difficult for wage earners to underreport income, employees have considerable leeway when it comes to claiming expenses that can be deducted from gross earnings to reduce net taxable income. The evidence concerning the drivers of the specific deterrence effect suggests that both volatility and sophistication matter. Business income differs from labour income also in terms of volatility allowing taxpayers to change re-

and as it will be shown shortly below, this behaviour is not a feature of the CIT compliance strategy of taxpayers in Rwanda, even though there is a marked, but not significant, reduction in the magnitude of the effect. The response to tax audits is also found in the literature to be sensitive to the audit outcome and so the average aggregate effect might mask considerable heterogeneity among taxpayers who, following the audit, have been verified (or not) additional tax to pay.[20] Generally, the literature provides evidence that the pro-deterrence impact of tax audits is driven entirely by audited taxpayers determined noncompliant while audited taxpayers determined compliant tend to show the opposite behaviour (see, for example, Gemmell and Ratto, 2012; Løyland and Øvrum, 2017; DeBacker et al., 2018a and Beer et al., 2020). While we also find evidence that the pro-deterrence effect of audits is due to a change in behaviour of audited taxpayers who have been identified noncompliant, the results do not provide conclusive evidence for audited taxpayers who have been identified as compliant: for the latter category of taxpayers the impact is not significant.

From a methodological point of view, the studies in the literature differ depending on the nature of the audit data analysed. With the risk of oversimplification, the empirical studies can be classified in two broad categories depending upon the data they use: the ones using random audits and the ones exploiting risk-based audits. Most studies in the literature predominantly examine the impact of random audits on future taxpayers' behaviour either exploiting tax administration operational data (see, for example, Gemmell and Ratto, 2012; Advani et al., 2021, and Best et al., 2021), or through randomized controlled trials (RCTs) on stratified samples (see, for example, Pomeranz, 2015, and Kleven et al., 2011). There are much fewer contributions in the literature evaluating the impact of operational risk-based audits (Løyland et al., 2019; Beer et al., 2020, and Erard et al., 2019) using tax administration operational data. No approach is better than the other and they both have merits. Studies employing random treatments, and in particular RCTs, have been heralded as being in the vanguard of the 'credibility revolution'[21]

ported income from year to year more easily. The implication of this is that the impact of audits on compliance are not as persistent for business income as it is for labour income. Sophisticated taxpayers, proxied by the ones with more experience in filing tax returns, are also less affected by tax enforcement (DeBacker et al., 2018a).

[20] And across margins as well: such as, for example, the sector the business belongs to, and whether it engages in exports/imports, the legal type of the business, and age of the business. Appendices B and C report robustness results across several margins (the sector businesses belong to, a finer classification of the business activity, indicators for the tax centre, late CIT reported and different sources of income and the VAT reporting behaviour of the firms) employed either as controls or additional matching variables. Other contributions have exploited different margins to shed light on potential heterogeneities in the deterrence effect of audits. For instance, Best et al. (2021) exploit the size and age of the firm, its industry, location, and position in the supply chain but they find no evidence of audit effect across those margins.

[21] See Slemrod, 2019.

regarding internal validity. The reason for this is because a random treatment provides two statistically identical clusters that differ in terms of one cluster receiving the policy treatment while the other does not thereby the causal effect of audits on treated taxpayers can be estimated through a DID approach. On the other hand, external validity of studies employing random audits could be more problematic because by using random audits—based on a population that extends beyond those taxpayers that are potential targets for audit—the estimated average compliance effect might not be sufficiently informative about the behavioural effects of taxpayers targeted for risk-based audits (see also Heckman and Smith, 1995).[22]

The approach based on the analysis of risk-targeted audits has the advantage of better external validity relative to random audits but its drawback being the internal validity. Indeed, while relying on risk-based audits is likely to produce a more generalizable evaluation of the change in behaviour of targeted taxpayers compared to RCTs, or other studies based on random audits, the estimation of the impact of enforcement policies on taxpayers' compliance might be more complicated since, by design, they focus on tax returns that are most suspected of noncompliance. Therefore, this approach entails the risk of overstating the magnitude of evasion. In order to overcome this limitation when estimating the causal impact of risk-based audits, the main estimation strategies entail the use of matching techniques (as in, for example, Li et al., 2019; Beer et al., 2020, and Erard et al., 2019), regression discontinuity design (as in, Løyland et al., 2019) or fine-tuned fixed effects models (as in DeBacker et al., 2015). Given the nature of RRA audits that target taxpayers based on their risk to evade and the nature of our data, this paper also contributes to this literature by extending the evidence on the impact of risk-based audits on deterring future noncompliance by combining different matching methods and a DID approach (Section 5 provides more details on the methodology).

# 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 and therefore any information transmitted to businesses regarding the quality of information held by RRA, and used in assessing their true tax liabilities, might be imperfect. Secondly, taxpayers rationally utilise any

[22]The policy relevance of the compliance effect estimated through random audits can thus be controversial and in some instances questionable, an issue discussed in, for example, Løyland et al. (2019), Beer et al. (2020), and Erard et al. (2019).

available information, which can be obtained from various sources[23] in order to infer the true probability of the being audited. Both of these elements affect the trade off between underreporting and being caught (and penalised) and not being caught and, therefore, the level of underreporting given true income.[24] 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.

To elaborate on the role of information obtained in taxpayers' updating their beliefs regarding the true probability of auditing, suppose that at time $\tau$ 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.[25] Notice that the prior might not reflect precisely the true probability distribution of $p$, an aspect that is captured by the $1/Var(p)$. As discussed in the preceding paragraph, business do obtain additional information during the audit process (and not only) regarding the probability of being audited and found underreporting. Denote the information received by the taxpayer[26] 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 might not be accurate, an aspect that is captured by the $1/E(Var(\tilde{p}|p))$, which gives the precision of the information received (if $1/E\left(Var(\tilde{p}|p)\right) \to 0$ $(\infty)$ then the information is inaccurate (accurate)).

Assuming now that the posterior density $z(\tilde{p}|p)$ and the prior density $g(p)$ give rise to a linear posterior density, the expected probability of being audited[27] at time $\tau$ is given by

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

[23]Such as the audit process, the sector the taxpayer belongs to, other audited businesses and the RRA itsel.

[24]The reasoning developed here follows that of Advani et al. (2021), DeBacker et al. (2015), and Best et al. (2021). Recent (experimental) evidence that tax audits may have differential effects on post-audit compliance is offered by Kasper and Alm (2022).

[25]This prior might reflect, for example, past experience and/or even any information provided by RRA in their annual reports regarding the (aggregate) likelihood of auditing.

[26]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.

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

and so it is a *weighted average*[28] of the taxpayer's prior mean of the probability of being audited $E(p)$ and the information obtained from the audit $\tilde{p}$, at time $\tau$, 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 taxpayers, $1/E\left(Var(\tilde{p}|p)\right)$, if and only if the expected prior, $E(p)$, is greater than the information received, $\tilde{p}$; otherwise it is increasing.

The point thus far is that information matters for the taxpayers and it informs the estimated probability of auditing which affects their decision to underreport.[29] If audits are not informative for the taxpayers then there is no reason for a taxpayer 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 taxpayer then a taxpayer 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.

To put the above into context consider the canonical model developed and analyzed by Allingham and Sandmo (1972) 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 $\pi$, 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

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

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

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,[30] 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)[31]

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

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 answer to this relies on how the expected probability of auditing, $E\left(p|\tilde{p}\right)$, is affected by the information obtained by the taxpayer. Clearly, 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 preceding discussion emphasises the role of information in taxpayers determining the likelihood of income underreporting being discovered by the Revenue Authority which, in turn, affects the future compliance behaviour of the taxpayer.

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

[30] A realistic assumption for Rwanda, given the size of large businesses there. A subscript denotes differentiation with respect to the argument given.

[31] Notice as 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).

# 4 Institutional setting and data

Rwanda is a country that embraces reforms and over the years the RRA has taken important steps to improve service delivery and enhance tax compliance.[32] Nevertheless, the RRA faces challenges that might hinder its performance and service delivery, including the COVID-19 outbreak and associated mitigation measures, that had a large impact on economic activity and a significant bearing on high domestic tax arrears, the low tax compliance culture by some taxpayers and insufficient allocated budget to clear all VAT refund backlog. Rwanda therefore provides an interesting framework for assessing tax audit strategies in developing countries.

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% of its tax revenues from CIT (the average of the CIT tax base for 2013-2018 is 17.24%) and VAT (the average of the VAT tax base for 2013-2018 is 33.06%). 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.

RRA classifies businesses as follows:[33] *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 a tax period; *Small*-businesses have a turnover between Rwf 12 million and Rwf 50 million (USD 55,750) in a tax period; and *Medium*-businesses have a turnover higher than Rwf 50 million in a tax period. The law also specifies that a business in order to be classified as *Large*-business should be notified by the RRA and registered accordingly as a large taxpayer. The analysis that follows employs the classification of businesses by their size as used by RRA in tax declaration records.

The CIT-real regime entails a standard tax rate of 30% on profit with some reductions available for specific groups.[34] Furthermore, small businesses can decide to benefit

[32] These include, among others, intensification of tax education and information programs, monitoring of non-filers and non-payers, enhancement of electronic billing machines, enhancement of the e-tax system and the local government tax management system, integration of e-tax system with the local government tax management system and further progress towards enhancement of an electronic single window system. Rwanda is a representative low-income country and one that has been implementing a series of successful economic and structural reforms. Rwanda's tax-to-GDP ratio, an aggregate proxy of fiscal capacity, is similar in both level and evolution over time to the average of African peer countries. In addition, the Rwanda's tax revenues structure seems to reproduce quite well that of an average low-income country, in particular when it comes to the share of CIT and VAT. According to OECD et al. (2020) data for the fiscal year 2017, these two tax bases represent almost half of the total tax revenues collected in Rwanda (46%), on average in Africa (48%), and in Latin America and Caribbean (44%), while the share is lower for the average of OECD countries (29%). The tax-to-GDP ratio in Rwanda has been steadily increasing from 10% in 1998 to 16.3% in 2018-2019, RRA (2019).

[33] This follows Article 2 of the Ministerial Order Nº002/13/10/TC.

[34] Articles 49 and 51 of the Law Nº016/2018 of 13/04/2018 guide reductions on CIT. These include, for example, facilitations for new businesses: newly listed companies on capital markets selling at least

from a simplified CIT-lump-sum tax regime having to pay a lump sum tax at the rate of 3% 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 payable 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 the tax payable—and withholding taxes. Businesses under the CIT-flat regime are required to file a considerably simplified form including their business income, which determine the correspondent flat amount to be paid, and tax credits claimed that coincide with the sum of quarterly instalments already paid. The analysis focuses on CTI and CIT payable reported by businesses as the two outcome variables. CTI reported across the tax regimes is the tax base upon which the corporate income tax is applied, while CIT payable is the tax payable by taxpayers net of any tax discounts claimed.[35]

Table 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% year on year.

This increase in the number of CIT filers is attributed to a number of factors. First,[36] RRA introduced the possibility for small and micro businesses to opt for the above-mentioned simplified tax regimes which have progressively incentivised 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, where some responsibilities of the division were assigned to other divisions allowing

20%, 30% or 40% of their shares to the public are taxed at a CIT rate of 28%, 25% or 20%, respectively, for a period of five years (see also PwC, 2019).

[35]More precisely, for CIT-real regime CTI corresponds to the total income obtained from different sources and calculated net of expenses, depreciation adjustments and deductions. When this calculation leads to a negative amount, the CTI 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, CTI reported corresponds to the total income declared from different sources while for taxpayers reporting under the CIT-flat regime CTI coincides with their business income. Given their favourable tax schedule, business reporting under the two simplified regimes may not reduce the CTI through deductions nor CIT through discounts.

[36]Through Law Nº28/2012 which modified and complemented Law Law Nº16/2005 on direct taxes on income.

Table 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.

the Registration and Block Management division to focus on following up on potential unregistered taxpayers. In early 2014, there was also the establishment of 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.[37]

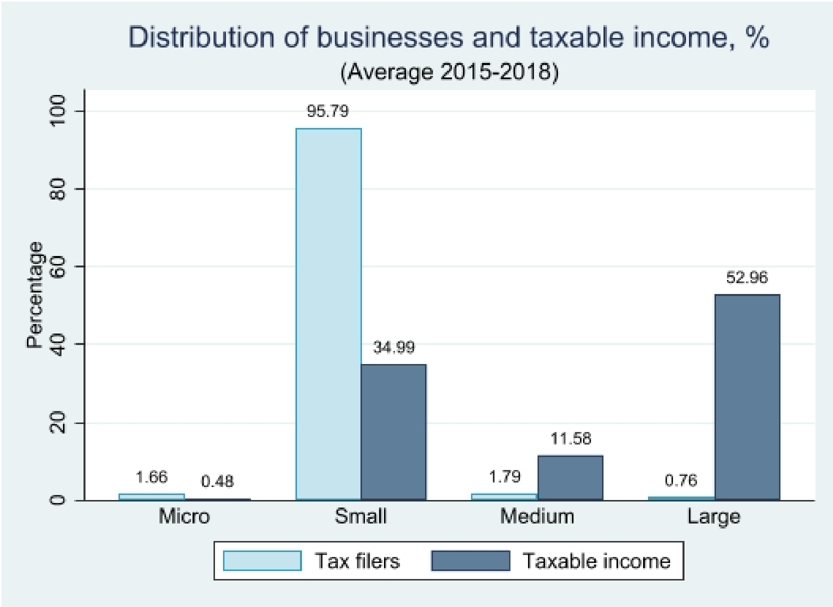
All data employed in this paper is at the taxpayer (business) level:[38] they include mostly financial variables used to calculate taxes (for example, total sales, taxable income, VAT refunds), as well as some taxpayer characteristics, such as sector and geographical location (at tax centre level). Figure 1 presents the average distribution of businesses reporting CTI by size across the period 2015-2018 together with the correspondent 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%), with Medium-businesses and Large-businesses consisting together of roughly 2.55% of the total while Micro-businesses represent 1.66% 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% of total CTI declared (52.96% large, and 11.58% medium). This pattern is even amplified in terms of revenues collected. Figure 2 shows that

[37] Most businesses make use of electronic billing machines, whose provision is described in Law N$^{o}$37/2012 of 09/11/2012 on the Code of VAT and the Ministerial Order no 002/13/10/TC of 31/07/2013. Article 18 specifies the three exempt categories from electronic billing machine registration. In particular, exempt category I refers to all non-VAT registered businesses being automatically qualified for exemption from the electronic billing machine. Exempt category II includes all VAT registered businesses where the scope of VAT sales is small compared to the company's total sales. Sales that sum up to 75% of the company's income derived from exempted services, are considered to belong with exempt category II and to be exempted they must apply for an exemption. Exempt category III includes all VAT registered businesses whose sales are only conducted during a limited portion of the year.

[38] See also Appendix A which provides additional information on the data set.

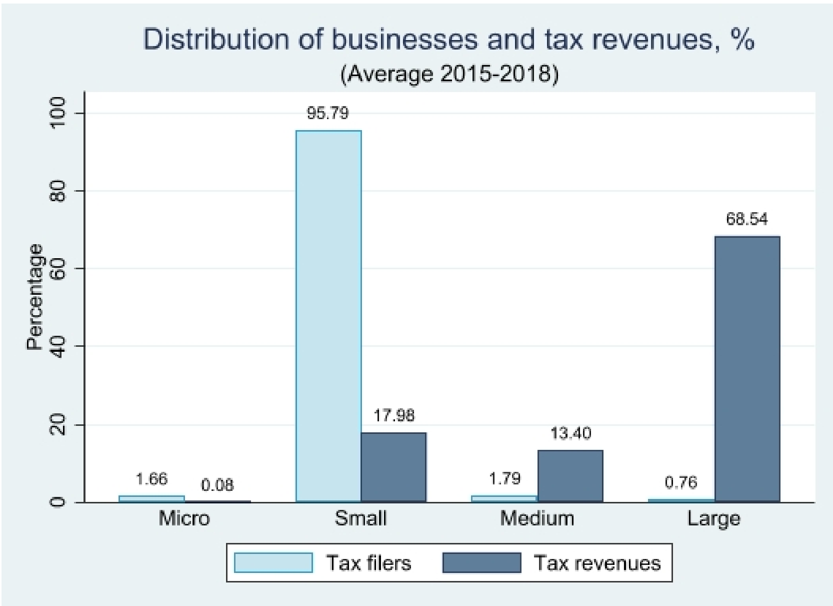
the main share of tax revenues is collected from large/medium businesses, which account together for 81.94% of total revenues collected (68.54% large, 13.40% medium).[39]

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




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

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



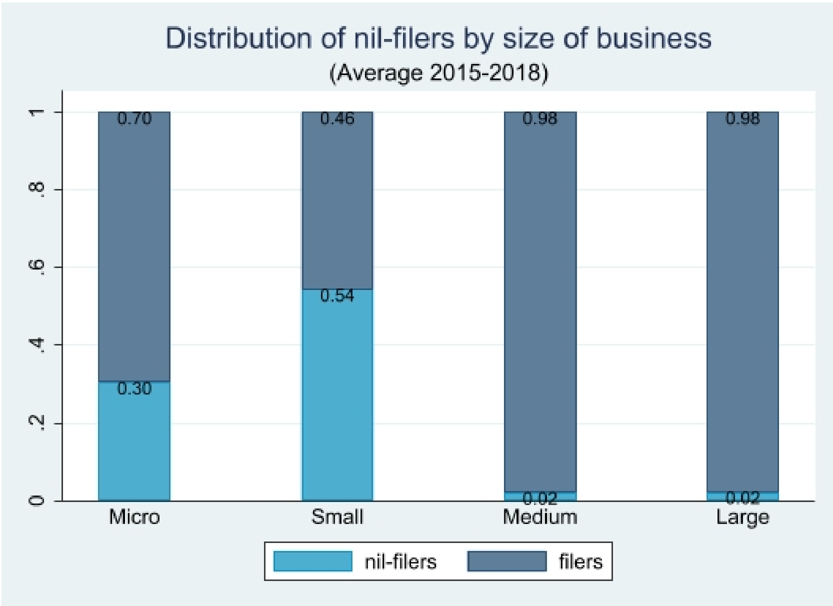

Note: Authors' calculations based on data provided by RRA.

A significant share of CIT filers are nil-filers (a nil-filer is a taxpayer who filed nil sales, input, output because the taxpayer has not operated any business activity during

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

a given tax period). 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% of micro-businesses are nil-filers, while 54% of small-businesses report zero on all fields of their tax declaration. The percentage of nil-filers is smaller for medium (2%) and large-businesses (2%).

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




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

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.[40] We have also been given access to the tax audit appeals[41] which is an important element of the audit process. Given the small number of appeals related to audits in the 2015 wave though, they do not seem to play a crucial role in determining the results of the analysis.

[40]Tax collection in Rwanda is managed by 30 tax centres appointed at the district level plus one ad-hoc tax centre for large companies based in Kigali (Kigali Large Taxpayer Office–LTO). The tax centres of the three districts which constitute the city of Kigali (Gasabo, Kicukiro and Nyarugenge) plus the Kigali LTO manage around 71% of CIT tax declarations. Tax auditing is decentralised, with the function being performed by the Offices, and follows taxpayer segmentation and not tax function.

[41]The taxpayer who is not satisfied with the contents of the tax assessment notice may appeal to the Commissioner General within thirty (30) days after receipt of the assessment notice. The appeal does not suspend the obligation to pay tax, interest and penalties. Upon written request by the taxpayer, the Commissioner General may suspend payment of the disputed amount of tax for the duration of the appeal. The tax audits considered are completed (and considered closed by the RRA) and there is outstanding appeal against them. See also footnote 44.

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 CIT declarations. The risk criteria utilise information that spans across tax bases.[42] The administrative data is retrieved from RRA systems which collect and store tax data from tax procedures followed by taxpayers. 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.

Tax audits in Rwanda are guided by Law Nº26/2019 on tax procedures which describe the tax audit process as well as the statute of limitations and the type of tax enforcement examinations undertaken by the RRA. Concerning the statute of limitations, RRA can audit a taxpayer for a period going back five years (this extended to ten years before the reform introduced with Law Nº26/2019). In general, however, 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 three types of audits:[43] desk audits, issue audits and comprehensive audits. Desk audits are conducted by RRA staff using 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 is not corresponding 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 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 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.[44]

[42] After each return has been filed, audit flags are generated based on the characteristics of the returns in a deterministic way.

[43] 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.

[44] This is described in particular in Law Nº26/2019 of 18/09/2019 on tax procedures. Following the completion of the audit, the RRA is required to issue a taxpayer with tax rectification note. The taxpayer is granted 30 days within which to respond. In case the taxpayer has not responded by that time, a final

During 2015 RRA performed 435 audits involving CTI reported, 389 (which accounts for 89.43% of all audits) of which uncovered CTI underreported by taxpayers with 217 leading to the application of tax fines (or 49.89%).

Table 2 presents summary statistics for the main outcome variables associated with the tax audits performed by RRA in 2015 and for all tax periods audited. The maximum amount of CTI 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 CTI 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 CTI 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 CTI declared by the taxpayer and *total audit outcome* as to include tax fines).

Table 2 reveals that audits contribute a substantial amount of tax revenues in terms of uncovered CTI underreported which amounts to 67.36% of the potential tax base audited in 2015 (71.47% including fines). Tables 3 and 4 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% of all underreporting detected while narrow-scope audits detect 8%.[45] In relative terms, and with respect to the correspondent self-assessment, comprehensive audits detect 60% of the potential tax base not including fines (62% including fines in the potential tax base). On the other hand, narrow-scope audits detect 72% of the potential tax base without fines (77% including fines).

To summarize, in order to assess the impact of tax audits on CIT filers' reporting

notice of assessment is issued. The taxpayer is allowed 30 days within which to appeal. Once an appeal is submitted to the Commissioner General, the RRA has 30 days within which to respond to the objection. This can be extended by another 30 days but not beyond this period. At this stage, the appeal is handled by the appeals committee, and the taxpayer and the taxpayer's agent are invited for a meeting to provide explanations. Once the final assessment is issued, any tax due is payable. However, the Commissioner General has powers to suspend the payment pending the determination of the appeal. There is a provision for resolving the dispute through an amicable settlement process. Taxpayers can opt for this approach while at the same time exploring the next stage of the appeal process. A taxpayer that disagrees with the response on the final assessment can appeal to the high court within 30 days.

[45] 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 2: 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 3: 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 4: 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.

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% of the 435 tax audits were comprehensive, 44.6% were desk and 18.39% were issue-oriented. 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.

- 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 Methodology

To estimate the impact of audits on future reporting behaviour we combine matching methods with a DID approach. Specifically the main objective of the analysis is to quantify the Average Treatment effect of audits on the reporting behaviour of Treated (that is, the audited) taxpayers ($ATT$). To this end, use will be made of two alternative outcome variables: CTI and CIT payable reported as defined in Section 4 and expressed in natural logarithms - ln hencheforth.[46] Denoting with $\Delta Y_i^1$ the change in taxpayer $i$'s reported outcome if the taxpayer has been audited and with $\Delta Y_i^0$ the analogous change if they were not audited and indicating the treatment assignment with the variable $D_i \in \{0, 1\}$, the $ATT$ is given by the difference

$$ATT = E\left[\Delta Y^1 | D = 1\right] - E\left[\Delta Y^0 | D = 1\right]. \tag{5}$$

The first term, $E\left[\Delta Y^1 | D = 1\right]$, can be estimated by employing the observed outcome variable reported of audited taxpayers before and after the audit process. The second term, given by $E\left[\Delta Y^0 | D = 1\right]$, gives the change in the reporting behaviour of audited taxpayers, had they not been audited, and it is not observable. In order to consistently estimate $ATT$, therefore, this counterfactual term needs to be proxied.[47]

Under the RRA risk-based audit framework the main challenge in the estimation of the counterfactual term of equation (5), and thus of the $ATT$, relates to the presence of a selection bias. Since audited taxpayers are chosen through risk-based assessment,

[46] To be more precise, we use the transformation $ln(x + 1)$ so to account for null values.

[47] In general the methods needed to estimate the counterfactual term in equation (5) differ depending on the type of audit data available.

the difference between treated and non-treated outcomes, in the absence of treatment, is in general not null (in the sense that $E\left[\Delta Y^0|D=1\right] - E\left[\Delta Y^0|D=0\right] \neq 0$) and thus the counterfactual term in equation (5) should be estimated by relying on identifying assumptions in order to address the selection problem.

To address this issue, and thus consistently estimate the $ATT$, the strategy is to employ matched-DID that consists of matching treated/audited taxpayers with similar taxpayers in a control group and apply a DID approach with that matched control group. Briefly, the basic idea behind matching is to pair each member of the treatment group with a set of observationally equivalent control group members. By holding the confounding factors constant, the difference between the outcome variable of audited (treated) taxpayers and matched controls (unaudited) is a direct estimate of the treatment effect that does not rely on any parametric assumptions (Guo and Fraser, 2015). More precisely, the key goal of matching is to prune (or, better, to weight) observations from the data so that there is better balance between the treated and control group, meaning that the empirical distributions of the relevant covariates in the groups are more similar. The distinction between different matching methods regards how similarity is synthetized and balance achieved (see Section 5.1 below for further details on this).

The analysis relies on the combination of several matching techniques. Coarsened Exact Matching (CEM), which will also be the main estimation strategy for the $ATT$, is employed as a data preprocessing step to stratify the sample on the relevant covariates. Propensity Score Matching (PSM) and Mahalanobis distance metric matching (MHD) estimators are then also employed on the CEM-stratified sample as robustness. We will also employ CEM-improved versions of the Inverse Probability of Treatment Weighting (IPTW) for the estimation of audit-type specific $ATT$. The initial stratification of data and the combination of different matching techniques help overcoming the limitations that may be encountered when these methods are applied in isolation. Section 5.1 below further elaborates on the estimation strategy presenting also details regarding each matching technique employed and introduces the sensitivity analyses performed. We thus estimate equation (5) with

$$\widehat{ATT} = \frac{1}{N^1} \sum_{i:D_i=1} \left[\Delta Y_i^1 - \hat{C}\right], \tag{6}$$

where $N^1$ is the number of treated taxpayers and $\hat{C}$ is the matching estimator providing an estimate of the counterfactual term $E\ \left[\Delta Y_i^0|D=1\right]$ by weighting the observation in the control group based on their similarities with the ones in the treatment group, that is

$$\hat{C} = \sum_{j:D_j=0} W\left(i,j\right) \Delta Y_j^0. \tag{7}$$

We employ several alternative specification of $\hat{C}$ based on the above-mentioned matching techniques, with the main distinction between them being in how the weights (given by $W(i,j)$) are calculated, something that is further discussed in more detail in Section 5.1.

To summarise, for each audited taxpayer each of our statistical matching approach selects one or more unaudited taxpayers as matched controls. These matched controls are then used to predict how audited taxpayers would have reported the outcome variables in future periods in the absence of tax auditing. A DID approach is then employed on the matched set of taxpayers to estimate $ATT$. Notice that any taxpayer audited between $01/04/t$ and $31/03/t+1$ is classified as 'treated' in wave $t-1$ in the senses 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 $t$).

The next section elaborates further on the identification strategy.

## 5.1 Estimation strategy

Several matching techniques are discussed in the literature (see, among others, Stuart, 2010; King et al., 2011; Imbens and Rubin, 2015; Guo and Fraser, 2015). In general, the use of matching methods for causal inference entails seeking a trade-off between maximizing balance on the relevent pre-treatment covariates between the treated and control units while keeping a reasonable matched sample size (King et al., 2011). The use of exact matching would ideally eliminate imbalance but at the cost of losing most of the available observations (Imai et al., 2008).

A whole set of approximate matching methods specifies a synthetic metric to assess similarity across treatment cohorts. Popular synthetic metrics are the estimated individual-specific probability of treatment assignment that is, the propensity score (Rosenbaum and Rubin, 1983) and the Mahalanobis distance between covariates. 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. CEM provides a stratification solution to approximate a fully blocked experimental design and possesses a set of powerful statistical properties that overcome these limitations. 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[48] as well as reducing model dependence,

[48]The main reason for this is that matching methods like PSM and MHD are designed to reduce the univariate imbalance in the mean of the propensity score (PSM) or each pre-treatment matching variable (MHD) across treatment cohorts but this does not necessarily achieve the desired reduction in multidimensional imbalance between the treated and control groups. With the purpose of overcoming

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 of the inference theory). CEM can also be used as a data preprocessing step to improve other matching methods since by applying those methods to CEM-stratified data one can combine their advantages with those provided by CEM and thus overcome their limitations.[49] For these reasons, we adopt CEM as the main estimation strategy for the estimation of the aggregate $ATT$ and provide, as robustness check, CEM-improved versions of PSM and MHD. For additional robustness, in Appendix B, we also augment all these models by incorporating regression specifications for the outcome variables controlling for potential residual imbalance across a comprehensive set of control variables beyond those employed in the stratification and matching. We return to this shortly below.

Moreover, the CEM procedure is extremely 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 in the expression of the counterfactual term $\hat{C}$ (in equation (7)), we analyse the unpruned units through a DID approach to finally estimate the $ATT$ given by equation (6).

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 taxpayers are required to file tax returns timely before treatment is applied in order to ensure that both

this limitation, Iacus et al. (2011) introduce a multivariate imbalance measure representing the distance between the multivariate empirical distributions of the treated and control units of the pre-treatment covariates.

[49]For instance, in our case by combining CEM and PSM it is possible to stratify the population on some variables particularly relevant for audit selection, while estimating the propensity score—the probability of being audited—on a more comprehensive set of variables to account for additional imbalance on these through a synthetic metric. As explained more in detail below, we adopt this approach among other robustnesses.

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 11,627 units. 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 62 of those in the treatment group and 344 in the control group. With this further restriction we are left with 11,221 observations to which we apply the CEM stratification procedure, further detailed below, to select the final matched sample of 5,881 units. Table 5 summarises the selection steps.

Table 5: Description of sample selection

Sample Selection

| Step | Description | Control Sample | % Δ | Audit Sample | % Δ | Total Sample | % Δ |
|---|---|---|---|---|---|---|---|
| 0 | Universe of CIT filers in 2015 | 28,619 | - | 435 | - | 29,174 | - |
| 1 | Drop outliers with effective tax rate >1 | 28,610 | 99.97% | 435 | 100.00% | 29,165 | 99.97% |
| 2 | Failure to file timely before treatment | 11,203 | 39.16% | 424 | 97.47% | 11,627 | 39.87% |
| 3 | Violation of (pre&post 2015) non-audit restrictions | 10,859 | 96.93% | 362 | 85.38% | 11,221 | 96.51% |
| 4 | Final matched sample after CEM | 5,577 | 51.36% | 304 | 83.98% | 5,881 | 52.41% |

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

The analysis employs as the main set of pre-treatment matching variables for CEM stratification, the synthetic Risk Index calculated based on the risk criteria provided by the RRA and the taxable income reported by taxpayers the three years before treatment is applied.[50] We choose these pre-treatment covariates because they provide us with enough

[50]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 we use relates to the former likelihood. What this means, in practical terms, is that there are businesses who declare the same corporate taxable income and have the same likelihood Risk Index, and some of those are audited (and are in the treated group) whereas some others are not (and are in the control group). Figure A.6 in Appendix A plots the probability of being audited across deciles of the Risk Index showing it is increasing. In particular, there is a correlation of 0.9113 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 CTI reported and Figure A.7 further elaborates on this by estimating the probability of being audited, by the combination of deciles of the two main matching variables, risk scores and CTI reported (for the year before treatment is applied). As discussed in Appendix A, all this points towards CTI reported picking-up in a non-linear way the role played by the unobserved impact-on-revenues likelihood.

information about the taxpayers' noncompliance riskiness—through the risk scores—and pre-treatment reporting behaviour over three years. This allows us to compare taxpayers with a common pre-treatment reporting trend which is likely to continue afterwards had the treatment not happened. 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. 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 as robustness and obtain results that are consistent with the results of the main analysis (see Appendix C).[51]

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 6. 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 6 shows the performance of the CEM procedure in reducing imbalance in our sample. The overall multivariate imbalance across pre-treatment covariates after CEM reduces to 45.9% of the initial imbalance while maintaining a considerably high share of treated taxpayers in the final matched sample (84%, see Table 5). The reduction in pre-treatment univariate imbalance is even more pronounced in particular for the Risk Index (for which imbalance reduces to just 25% of the initial value) indicating that homogeneity across treatment cohorts increases significantly in these covariates as a result of the CEM process. This is visually confirmed in Figure 4 that plots the distribution of pre-treatment covariates used for stratification before and after the CEM procedure. Similarly, Figure 5 provides a bidimensional visual confirmation of the multivatiate balance enhancement by showing that sufficient matches can be found simultaneously on both risk scores and

[51] More precisely, in addition to the variables employed in the main stratification set, broader sets of variables employed as robustness also include: information on taxpayers' adoption of EBM (exact matching), the amount of losses reported, a broad definition of the sector of activity (exact matching), VAT paid on inputs, tax centre indicators (exact matching) and lags of these variables among others. CEM-matching on these broader sets of variables leads to matching solutions that may lead to less generalizable results due to the (sometimes important) reduction in the size of the matched sample (for a more detailed discussion see Appendix C). Nevertheless, the results of this robustness analysis are consistent with the results presented in the main text both qualitatively and quantitatively.

taxable income reported in each of the three years before the start of the audit process. As further evidence of these balance improvements, Table 7 reports the mean difference across treatment cohorts in the matched sample as well as distributional differences at different percentiles both for the matching variables and other variables available for all the taxpayers in the matched sample.[52] As a result of the CEM procedure, the difference in the distribution of the matching variables across treatment cohorts is generally very low and never statistically distinguishable from zero. This is mostly the case also for the other variables available for all the matched observations.

As already discussed, exact matching on the business activity as additional variable in the stratification process is costly in terms of significant reduction in the sample size and it is not always feasible due to missing data. However, Figure 6 shows that the distribution of firms' business activity—as defined by the ISIC section—is fairly similar across treatment cohorts in our matched sample. While residual imbalance remains on some of the covariates not involved in the CEM stratification, we take care of this by augmenting all the matching models presented in the main text incorporating regression specifications for the outcome variables.[53] This allows us to control for a comprehensive set of variables that may influence the evolution of our outcome variables and thus confound the impact of the audit process (see Appendix B for more details). Results of this specifications are presented and discussed in Appendix B and corroborate the results presented in the main text.

As alternative specification to equation (7) we employ alternative matching methods on the same CEM-stratified sample. Namely, we use the Kernel PSM estimator, the

[52]As discussed in Section 4, the existence of different tax regimes entails that businesses have to provide different amount of information across regimes and more generally firms tend to fill different fields leaving others in blank. Here we report the fields commonly reported before treatment by all firms in the matched sample.

[53]We do so by performing weighted regression models (employing weights from our baseline models) as well as double-robust regression adjustment models for more robust inference (see, for example, Cattaneo, 2010; Wooldridge, 2002, 2007, and Imbens, 2004). Our weighted regression models control for a comprehensive set of variables: the risk score calculated for each of the three years before treatment, the taxable income reported in 2014 and 2013, the VAT paid on inputs reported each of the three years before treatment, indicator variables for the tax centre, the sector of activity and the section of activity (according to the International Standard Industrial Classification, ISIC), dummies for diverse type of income reported each of the three years before treatment and a dummy for lately reported CIT tax return in the year of treatment. A dummy variable indicating the adoption of EBMs before the treatment period is also included. The regression adjustment outcome models employ two alternative sets of covariates. Set I includes an indicator variable for the adoption of EBMs before the treatment period, the risk scores for the latest two pre-treatment years, reported taxable income declared in the year before treatment and a dummy for the sector of activity. Set II adds dummies for diverse type of income reported each of the three years before treatment, a dummy for CIT tax return reported after the deadline during the year of the audit process and a dummy identifying the four tax centers in Kigali (see Appendix B for more details). The results from these analyses corroborate our main results and they are presented and discussed in Appendix B.

Table 6: Imbalance pre and post CEM matching

Panel A: Overall imbalance, Multivariate $L_1$

| | |
|---|---|
| $L_1$ statistic pre CEM: | 0.61 |
| $L_1$ statistic post CEM: | 0.28 |

Panel B: Univariate imbalance

| | $L_1$ pre CEM | $L_1$ post CEM |
|---|---|---|
| Risk Index | 0.48 | 0.12 |
| CTI 2013 | 0.14 | 0.08 |
| CTI 2014 | 0.19 | 0.07 |
| CTI 2015 | 0.18 | 0.06 |

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

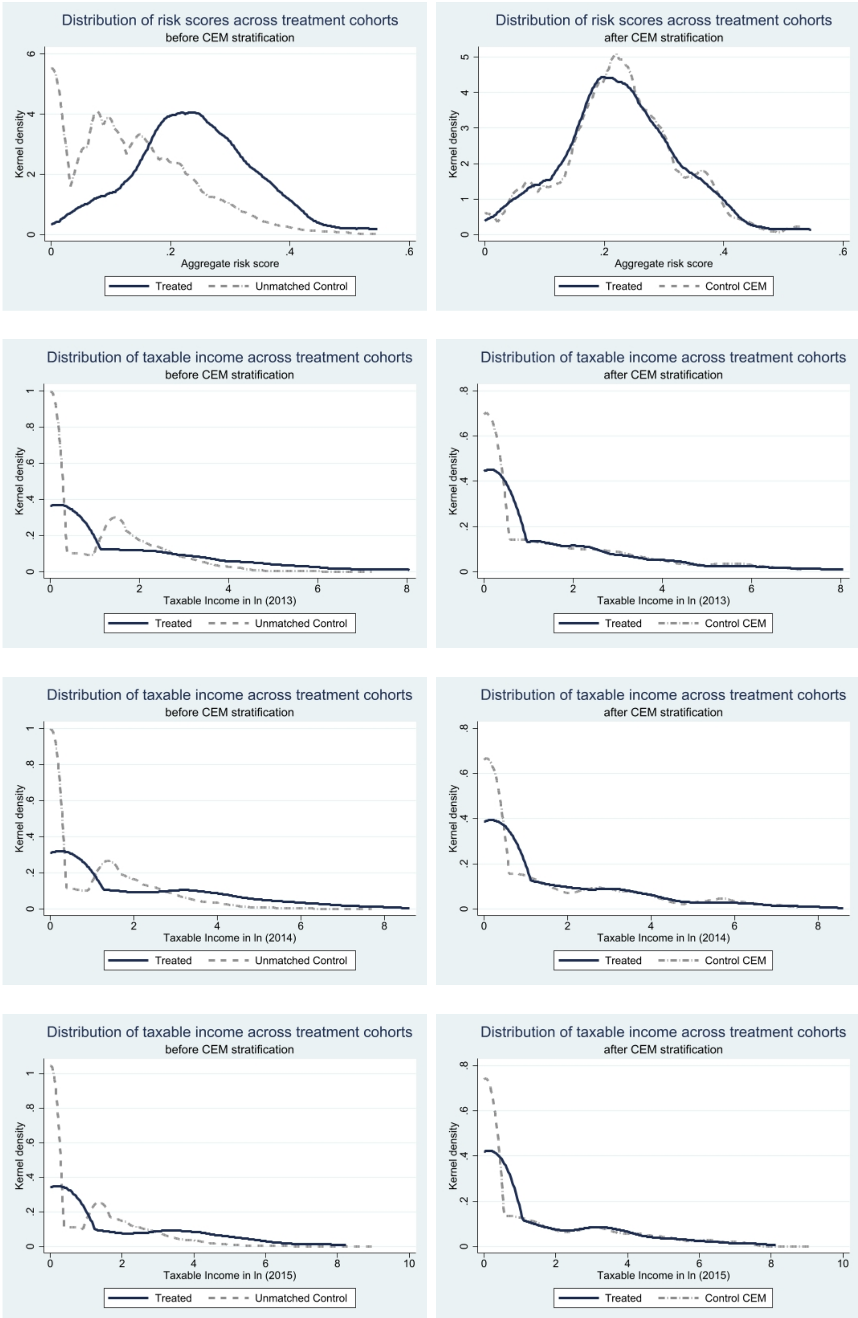
Table 7: Balance in the matched sample

| | Mean | 0.25 | Median | 0.75 |
|---|---|---|---|---|
| Matched variables | | | | |
| Risk Index | -0.001 | 0.000 | -0.009 | -0.003 |
| | (0.007) | (0.007) | (0.006) | (0.008) |
| CTI 2013 (ln) | 0.024 | 0.000 | 0.012 | -0.142 |
| | (0.130) | (0.028) | (0.037) | (0.258) |
| CTI 2014 (ln) | 0.037 | 0.000 | 0.045 | 0.083 |
| | (0.138) | (0.030) | (0.053) | (0.324) |
| CTI 2015 (ln) | 0.013 | 0.000 | 0.000 | 0.147 |
| | (0.138) | (0.026) | (0.034) | (0.451) |
| Other variables | | | | |
| Losses reported 2015 | 262.645 | 0.000 | 0.000 | 12.277 |
| | (159.988) | (0.006) | (0.008) | (8.694) |
| CIT payable 2015 (ln) | 0.036 | 0.000 | 0.000 | 0.047 |
| | (0.105) | (0.009) | (0.012) | (0.218) |
| Balance Due 2015 (ln) | 0.007 | 0.000 | 0.000 | 0.024 |
| | (0.078) | (0.001) | (0.002) | (0.069) |
| VAT paid on input 2015 (ln) | 2.479*** | 0.000 | 2.879*** | 1.536*** |
| | (0.332) | (0.164) | (0.559) | (0.261) |
| EBM adoption | -0.005 | 0.000 | 0.000 | 0.000 |
| | (0.028) | (0.001) | (0.001) | (0.380) |
| Observations | 5881 | 5881 | 5881 | 5881 |

Note: Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Coefficients denote the difference in the correspondent metric between audited and unaudited taxpayers in the matched sample. Inference is obtained through weighted regressions and weighted quantile regressions using weights from the CEM stratification.

Kernel MHD estimator and the Nearest Neighbour MHD estimator. Based on a set of

Figure 4: Univariate imbalance reduction (CEM)

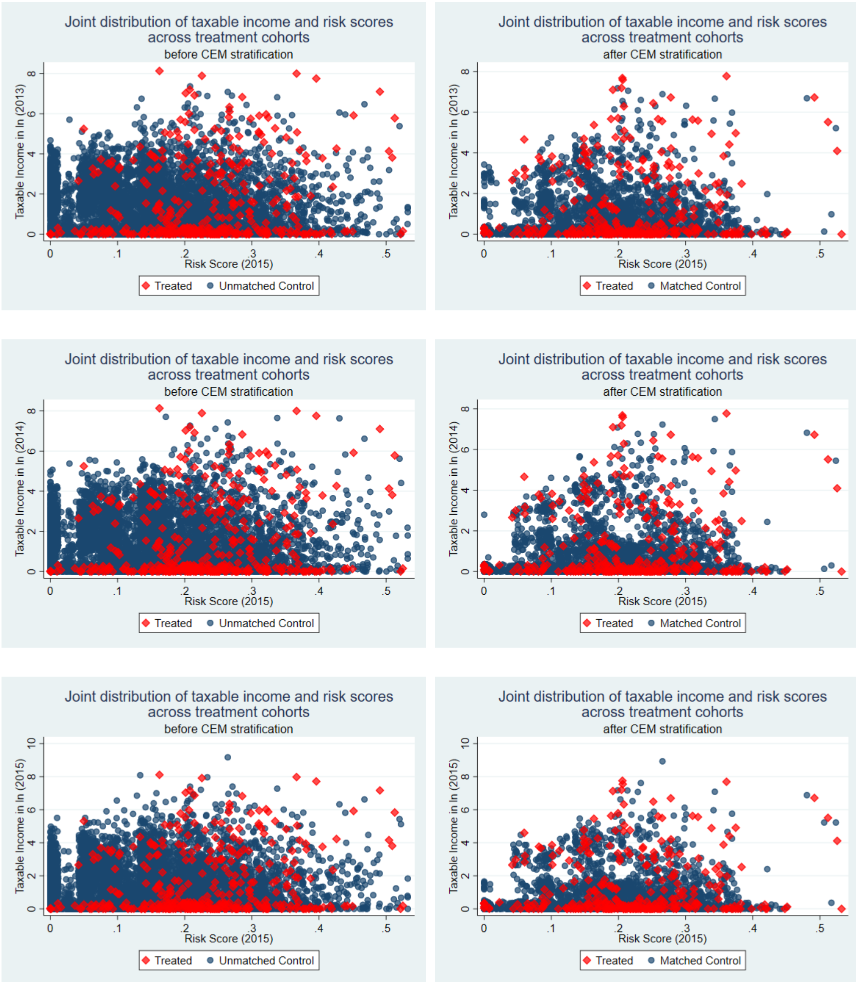



Note: Authors' calculations based on data provided by RRA.

pre-treatment covariates,[54] Kernel PSM estimates the propensity score through a discrete choice model (in our case probit) where the dependent variable reflects assignment to

[54]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 nine variables, namely: the synthetic Risk Index calculated for each of the three years before treatment is applied, the taxable income and the VAT paid on inputs reported by taxpayers for each of the three years before treatment is applied.

Figure 5: Joint distribution of Risk Index and CTI before and after CEM



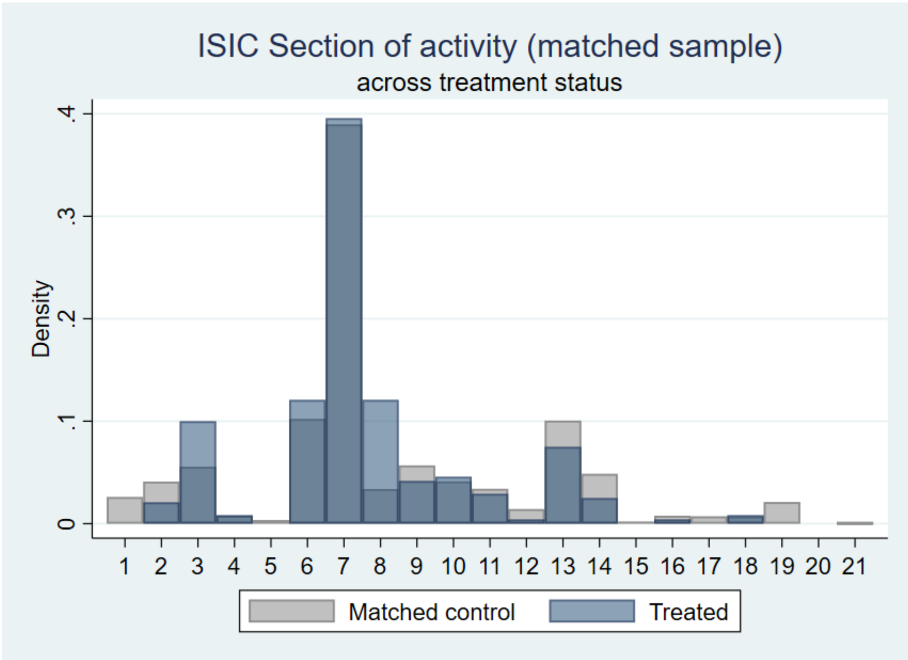

Note: Authors' calculations based on data provided by RRA.

audit treatment and subsequently estimates the counterfactual term as a simple weighted average of the outcomes with weights reflecting the similarity across participants.[55] More precisely, the Kernel PSM estimator assigns weights in equation (7) 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

[55] 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.

Figure 6: Distribution of firms by ISIC section of activity across treatment status (matched sample)




Note: Authors' calculations based on RRA data. ISIC sections of activities are coded between 1 and 21 in the following way: 1 Agriculture, forestry and fishing; 2 Mining and quarrying; 3 Manufacturing; 4 Electricity, gas, steam and air conditioning supply; 5 Water supply; sewerage, waste management and remediation activities; 6 Construction; 7 Wholesale and retail trade, repair of motor vehicles and motorcycles; 8 Transportation and storage; 9 Accommodation and food service activities; 10 Information and communication; 11 Financial and insurance activities; 12 Real estate activities; 13 Professional, scientific and technical activities; 14 Administrative and support service activities; 15 Public administration and defence; compulsory social security; 16 Education; 17 Human health and social work activities; 18 Arts, entertainment and recreation; 19 Other service activities; 20 Activities of households as employers, undifferentiated goods- and services-producing activities of households for own use; 21 Activities of extraterritorial organizations and bodies.

but employs the Mahalanobis distance metric for pre-treatment covariates,[56] instead of the distance between propensity scores $p_j - p_i$, to define the similarity between units $i$ and $j$ and to calculate weights in equation (7).

Finally, the Nearest Neighbour MHD estimator also utilizes the Mahalanobis distance metric to 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 member $j$ with the lowest level of the Mahalanobis distance for the relevant covariates and 0 to the others.

Thus far we have discussed the methodology to address the issue at the heart of this paper: 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 Stuart, 2010, and Erard et al., 2019) on the CEM-matched sample.[57] IPTW estimates a type-specific counterfactual term $\hat{C}$ for any different type of treatment (equation (7)) estimating a specific $ATT$ for any treatment (equation (6)). 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 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 ($p_i^{na}$). Formally, the estimate of equation (7) for the comprehensive type of audit is given by the expression $\hat{C}^c = \frac{1}{N^0}\sum_{j:D_j=0}\frac{p_j^c}{p_j^{na}}\Delta Y_j^0$ and the analogous expression of the estimated counterfactual outcome for desk-based audits is provided by $\hat{C}^d = \frac{1}{N^0}\sum_{j:D_j=0}\frac{p_j^d}{p_j^{na}}\Delta Y_j^0$ .

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). Standard bootstrapping methods are usually applied in this context but they cannot be generally justified. Indeed, while in some cases they provide valid inference because the estimators are asymptotically linear (as, for example, for Kernel-based methods), in some other cases they result

[56] 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 $X$.

[57] As additional robustness checks to this analysis we also perform double-robust regression adjustment versions of the IPTW model (IPTW-RA) simultaneously estimating an outcome model (see, for example, Cattaneo, 2010; Wooldridge, 2002, 2007). The results obtained validate our main results and are discussed in Appendix B.

in biased estimates for the standard errors.[58] For example, Abadie and Imbens (2008) have shown that standard bootstrapping is not asymptotically valid for Nearest Neighbour estimators with a fixed number of matches and Abadie and Imbens (2006) provide a valid analytical alternative to bootstrapping for this estimator. 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. Finally, 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. Given these premises, we provide inference by reporting robust standard errors (clustered by tax center) for the CEM and IPTW estimators, bootstrapped standard errors (based on 200 replications) for Kernel PSM and Kernel MHD, and the heteroskedasticity-consistent standard errors proposed by Abadie and Imbens (2006, 2008) for Nearest Neighbour MHD.[59]

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 $ATT$. In appendix we present further sensitivity analysis which validates the results presented in the main text and the methodology used.[60]

## 6.1 Aggregate $ATT$

A crucial assumption for any DID analysis is the existence of a common previous trend in the outcome variable at the time of the treatment (Meyer, 1995). Under this assumption,

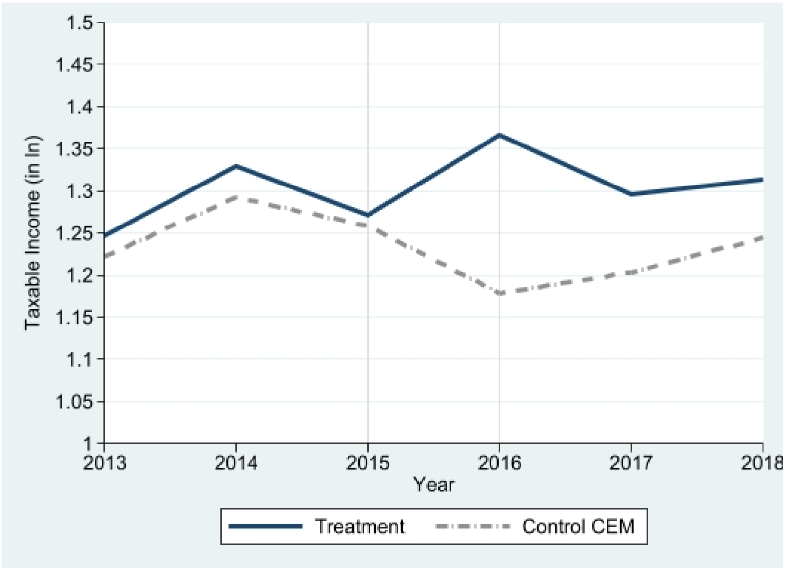
[58] The literature suggests that while standard bootstrapping is suitable for 'smoother' treatment effect estimators as, for example, Kernel-based methods, wild bootstrapping represents a theoretically justified bootstrap procedure for estimators with a fixed number of matches as, for example, Nearest Neighbour (see Bodory et al., 2020).

[59] Additionally, we also report alternative standard errors for any specification. In particular we report bootstrapped standard errors clustered at tax center level (based on 500 replications) for CEM, bootstrapped standard errors (based on 500 replications) for Kernel PSM, Kernel MHD and IPTW and wild bootstrapped standard errors (based on 500 replications) for Nearest Neighbour MHD. Finally, for any specifications, we also report stratified bootstrapped standard errors based on 500 replications and the strata resulting from the CEM pre-processing procedure.

[60] In particular, as already mentioned, Appendix B presents the results of our regression adjustment models and Appendix C reports the results of the estimation of our models obtained by applying more inclusive sets of covariates for the CEM stratification. Moreover, Appendix D performs the estimation of the audit-type-specific $ATT$ separately for different groups of businesses based on their probability of being audited. Finally, Appendix E provides further validations of our estimation strategy by estimating the dynamic effects of audits performed in 2013 wave (Subsection E.1), by testing the sensitivity of our main results to random subsampling and subsampling targeting outliers (Subsection E.2), and by providing a placebo impact estimate for a cluster of taxpayers audited after the 2016 tax return was reported treating them as they were audited in the 2015 wave (Subsection E.3).

one should observe a similar pattern in the evolution of the reporting behaviour of audited and unaudited taxpayers before treatment. Figure 7 presents the estimated $ATT$ obtained using CEM, showing the evolution of the ln-converted CTI reported across treatment cohorts. Noticeably, as a result of the CEM process, the ln-converted CTI reported presents not only a very similar trend in the periods before treatment but it is also comparable in levels across treatment cohorts. After treatment, the estimates clearly indicate a positive effect of audits on subsequent tax reporting behaviour of audited taxpayers. A similar result is obtained for CIT payable reported expressed in ln as shown in Figure 8.

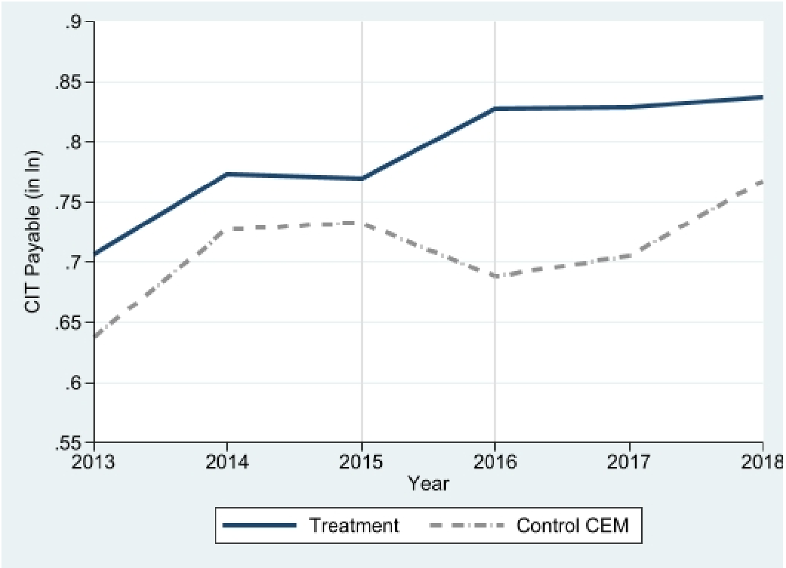
Figure 7: $ATT$ of Audits on audited taxpayers (CEM): CTI reported (in ln)




Note: This plot corresponds to results reported in Table 8, first row (models 1 to 3).

Table 8 reports the aggregate estimation of the $ATT$ for the audit wave of 2015. Given that the dependent variables are expressed in ln, the results can be interpreted as semi-elasticities. Columns (1) and (4) report the estimated $ATT$, respectively, on reported CTI and CIT payable reported one year after the audit. The CEM estimation strategy indicates that audited taxpayers tend to report about 17.5% more CTI the year after auditing, relative to similar matched taxpayers who have not been audited. This translates into around 10.3% more CIT payable reported compared to the control group. By employing the alternative matching techniques as robustness checks, the results are qualitatively and quantitatively very similar leading to relatively low model uncertainty. Indeed, regarding CTI reported (CIT payable reported) the largest estimated impact obtained one year after treatment using Nearest Neighbour MHD matching is 29.7% (14.7%) while the lowest obtained using kernel PSM is 14.8% (10.3% using CEM) and the ratio between these two estimates is 1.69 (1.44). All these results are robust across alternative

Figure 8: *ATT* of Audits on audited taxpayers (CEM): CIT Payable reported (in ln)




Note: This plot corresponds to results reported in Table 8, first row (models 4 to 6).

inference methods and provide evidence of significant and sizable pro-deterrence effect of audits on corporate income reporting behaviour one year after the audit process has started. On average across all the methods, treated taxpayers tend to report about 20.7% (12.3%) more CTI (CIT) one year after the end of the audit process when compared to the control group. Table 8 also reports the estimates two and three years after the audit (Columns 2-3 and 5-6). The effect is lower in magnitude when time passes but not statistically significant.

As already discussed in Section 2, the existing literature has pointed out that the impact of audits might be different depending on the outcome of the tax inspection. In order to verify whether this is also the case for the framework under analysis, we replicate the analysis by differentiating the sample based on the outcome of the audit process.[61] Table 9 presents the results of this analysis which suggest the main effect is driven by audited taxpayers determined noncompliant while it is inconclusive for the audited taxpayers determined compliant. Indeed, while for the cluster of taxpayers determined noncompliant the results are significant and comparable, both in sign and size to the ones presented in Table 8, we cannot draw a clear picture for the very residual sub-sample of taxpayers determined compliant (10% of the treatment group).

[61] This is done by sub-sampling based on audit outcome before the selection steps and the CEM stratification are applied to the data.

Table 8: Main Results – Aggregate $ATT$

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Matching estimator | (1) | (2) | (3) | (4) | (5) | (6) |
| CEM | 0.175 | 0.080 | 0.056 | 0.103 | 0.087 | 0.033 |
| | (0.023)*** | (0.147) | (0.111) | (0.017)*** | (0.107) | (0.081) |
| | (0.033)*** | (0.205) | (0.136) | (0.028)*** | (0.147) | (0.098) |
| | (0.085)** | (0.109) | (0.118) | (0.061)* | (0.079) | (0.083) |
| Kernel - MHD | 0.208 | 0.003 | 0.025 | 0.124 | 0.030 | 0.012 |
| | (0.023)*** | (0.147) | (0.111) | (0.017)** | (0.107) | (0.081) |
| | (0.084)** | (0.100) | (0.097) | (0.057)** | (0.068) | (0.069) |
| | (0.072)*** | (0.091) | (0.088) | (0.047)*** | (0.065) | (0.062) |
| Kernel - PSM | 0.148 | -0.074 | -0.145 | 0.119 | 0.023 | -0.059 |
| | (0.081)* | (0.107) | (0.117) | (0.059)** | (0.073) | (0.081) |
| | (0.085)* | (0.103) | (0.114) | (0.059)** | (0.071) | (0.083) |
| | (0.080)* | (0.102) | (0.103) | (0.054)** | (0.070) | (0.083) |
| Nearest Neighbour | 0.297 | 0.125 | 0.195 | 0.147 | 0.079 | 0.097 |
| | (0.099)*** | (0.120) | (0.143) | (0.072)** | (0.084) | (0.096) |
| | (0.095)*** | (0.218) | (0.198) | (0.072)** | (0.164) | (0.145) |
| | (0.115)*** | (0.143) | (0.162) | (0.080)* | (0.102) | (0.113) |

Note: Alternative standard errors are reported in parentheses for any specification. In particular, robust standard errors (clustered by tax center), bootstrapped standard errors (clustered by tax center) based on 500 replications, and stratified bootstrapped standard errors based on 500 replications are reported for CEM estimator; bootstrapped standard errors based on 200, 500 replications and stratified bootstrapped standard errors based on 500 replications are reported for Kernel - MHD and Kernel - PSM estimators; and heteroskedasticity-consistent analytical standard errors proposed by Abadie and Imbens (2006, 2008), wild bootstrapped standard errors based on 500 replications and stratified bootstrapped standard errors based on 500 replications are reported for Nearest Neighbour estimator; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.1.

Table 9: $ATT$ by audit outcome - Determined Noncompliant vs. Determined Compliant

| | Determined Noncompliant | | | | | | Determined Compliant | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dep. Variable | CTI reported | | | CIT payable reported | | | CTI reported | | | CIT payable reported | | |
| After audit | I | II | III | I | II | III | I | II | III | I | II | III |
| Estimator | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| CEM | 0.166 | 0.086 | 0.049 | 0.097 | 0.105 | 0.036 | 0.248 | 0.039 | 0.123 | 0.152 | -0.051 | 0.006 |
| | (0.034)*** | (0.172) | (0.123) | (0.020)*** | (0.123) | (0.089) | (0.151) | (0.260) | (0.217) | (0.120) | (0.168) | (0.125) |
| | (0.040)*** | (0.229) | (0.138) | (0.025)*** | (0.162) | (0.100) | (0.176) | (0.303) | (0.277) | (0.141) | (0.197) | (0.169) |
| | (0.097)* | (0.121) | (0.129) | (0.068) | (0.188) | (0.093) | (0.116)** | (0.212) | (0.245) | (0.079)* | (0.137) | (0.148) |
| Kernel - MHD | 0.212 | 0.022 | 0.033 | 0.128 | 0.058 | 0.023 | 0.089 | -0.043 | -0.072 | -0.008 | -0.089 | -0.015 |
| | (0.080)*** | (0.101) | (0.106) | (0.055)** | (0.069) | (0.073) | (0.174) | (0.281) | (0.199) | (0.093) | (0.164) | (0.120) |
| | (0.090)** | (0.108) | (0.106) | (0.061)** | (0.075) | (0.075) | (0.178) | (0.267) | (0.209) | (0.099) | (0.164) | (0.117) |
| | (0.085)** | (0.100) | (0.100) | (0.056)** | (0.072) | (0.069) | (0.168) | (0.257) | (0.200) | (0.091) | (0.150) | (0.115) |
| Kernel - PSM | 0.152 | -0.042 | -0.098 | 0.124 | 0.055 | -0.023 | 0.058 | -0.351 | -0.372 | 0.029 | -0.260 | -0.248 |
| | (0.086)* | (0.110) | (0.122) | (0.060)** | (0.078) | (0.087) | (0.155) | (0.283) | (0.336) | (0.109) | (0.179) | (0.216) |
| | (0.093) | (0.116) | (0.129) | (0.064)* | (0.082) | (0.091) | (0.164) | (0.298) | (0.346) | (0.109) | (0.192) | (0.222) |
| | (0.086)* | (0.105) | (0.112) | (0.065)* | (0.075) | (0.077) | (0.139) | (0.266) | (0.302) | (0.093) | (0.170) | (0.206) |
| Nearest | 0.320 | 0.182 | 0.206 | 0.184 | 0.141 | 0.087 | 0.351 | -0.009 | 0.14 | 0.151 | -0.083 | 0.054 |
| Neighbour | (0.143)** | (0.146) | (0.188) | (0.102)* | (0.118) | (0.140) | (0.207)* | (0.309) | (0.315) | (0.141) | (0.186) | (0.196) |
| | (0.144)** | (0.249) | (0.156) | (0.090)** | (0.184) | (0.083) | (0.260) | (0.216) | (0.251) | (0.151) | (0.119) | (0.173) |
| | (0.129)** | (0.156) | (0.174) | (0.093)** | (0.110) | (0.118) | (0.212)* | (0.311) | (0.339) | (0.122) | (0.184) | (0.202) |

Note: Alternative standard errors are reported in parentheses for any specification. In particular, robust standard errors (clustered by tax center), bootstrapped standard errors (clustered by tax center) based on 500 replications, and stratified bootstrapped standard errors based on 500 replications are reported for CEM estimator; bootstrapped standard errors based on 200, 500 replications and stratified bootstrapped standard errors based on 500 replications are reported for Kernel - MHD and Kernel - PSM estimators; and heteroskedasticity-consistent analytical standard errors proposed by Abadie and Imbens (2006, 2008), wild bootstrapped standard errors based on 500 replications and stratified bootstrapped standard errors based on 500 replications are reported for Nearest Neighbour estimator; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.1.

## 6.2 Audit-type specific $ATT$

As already discussed in Section 4, narrow-scope audits on the one hand and comprehensive audits on the other are considerably different types of tax enforcement examinations.[62] Table 10 presents the results of the estimation of the $ATT$ by audit type conducted via IPTW.

Table 10: Main Results – $ATT$ by audit type

| Dep. Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after audit | I | II | III | I | II | III |
| Type of Audit | (1) | (2) | (3) | (4) | (5) | (6) |
| Comprehensive | 0.285 | 0.130 | -0.040 | 0.246 | 0.136 | 0.030 |
| | (0.162)* | (0.228) | (0.241) | (0.128)* | (0.185) | (0.161) |
| | (0.173)* | (0.216) | (0.193) | (0.135)* | (0.173) | (0.139) |
| | (0.173)* | (0.205) | (0.168) | (0.135)* | (0.162) | (0.132) |
| Narrow-scope | 0.020 | -0.235 | -0.170 | 0.006 | -0.095 | -0.078 |
| | (0.030) | (0.066)*** | (0.046)*** | (0.026) | (0.047)** | (0.042)* |
| | (0.074) | (0.103)** | (0.107) | (0.041) | (0.062) | (0.063) |
| | (0.065) | (0.088)*** | (0.094)* | (0.036) | (0.053)* | (0.058) |

Note: Alternative standard errors are reported in parentheses. In particular we report robust standard errors (clustered by tax center), bootstrapped standard errors based on 500 replications and stratified bootstrapped standard errors based on 500 replications; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

According to this analysis, comprehensive audits drive the aggregate results presented in Table 8 being the only type of audit to present a pro-deterrence effect after one year from the start of the audit process and a trend similar to the aggregate $ATT$ in subsequent years. Interestingly, narrow-scope audits present a non-significant impact after one year but seem to have a counter-deterrence effect starting from the second year. More precisely, comprehensive audits lead to an average increase of about 28.5% (24.6%) in CTI (CIT payable) reported by audited taxpayers after receiving this type of audit when compared to the control group. Less intense type of audits tend to have a non-significant effect the first year after the audit and start to have a counter-deterrent effect from the second year after the audit leading to a reduction of 23.5% (9.5%) in CTI (CIT payable) reported by businesses that experienced this kind of audit.[63] For narrow-scope audits, the

[62] 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.

[63] A result which is consistent with the experimental evidence provided in Kasper and Alm (2022). Appendix D further elaborates on these results by performing the estimation of the audit-type specific $ATT$ separately for different groups of businesses and showing that the results are not uniform across types of businesses.

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 a complex issue that might relate also to the perceived fairness of auditing, broadly interpreted to include the penalties imposed on the underreported income as well as the process of settling tax disputes.

# 7 Concluding remarks

Improving administration of tax systems, and in particular its tax enforcement dimension, is undoubtedly a major challenge for Revenue Authorities across the world, and in particular so for developing countries, IMF (2015). Important for domestic revenue mobilization is understanding the role of operational audits and particularly their impact on deterring future noncompliance. This is an important issue and one that is directly related to the design and effectiveness of tax auditing and capacity building in tax administration.

By using available data on CIT, VAT and tax audits performed during the 2015 audit wave, this paper has investigated the role of tax enforcement in Rwanda, a developing country which has embraced reforms since 2000. The analysis has identified that tax audits in Rwanda deliver sizeable pro-deterrence effects on future corporate income reporting but there are also margins that can improve the performance of its tax enforcement policy.

The evidence suggests that there is in general a sizable pro-deterrence effect of tax audits on CTI and CIT payable reported by audited businesses one year after the start of the audit process, with on average audited businesses reporting about 20.7% more CTI the year after the beginning of the audit process, relative to similar matched businesses who have not been audited. This converts into about 12.3% more CIT payable reported compared to the businesses in the control group and corresponds to approximately 2.8% of total CIT payable declared by all incorporated businesses that year. The effect estimated is lower in magnitude in subsequent years, but not statistically significant, and appears to be completely driven by audited businesses determined noncompliant. The results are robust across different approaches.

RRA employs different types of tax examinations characterized by diverse degrees of intensity. 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 US 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. What this points to is that, from a policy perspective, Revenue Authorities should pay close attention to the evaluation of tax audits and their types.

The analysis also suggests avenues for future research. It is not entirely clear, for example, what drives the differential impact of the type of audits on compliance behaviour. As tax authorities pay increasing attention to narrow-issues audits, understanding why such audits could make businesses less compliant is important from an audit strategy perspective. The impact of audits on compliance depends not only on the reporting behaviour of businesses but also on the enforcement behaviour of tax authorities, and disentangling the two requires attention.

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 the meta analysis in 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 discussion between comprehensive versus narrow-scope audits, it remains to be seen. The analysis has only considered the 2015 CIT audit wave.[64] A natural question then to ask is: do past tax audits affect future CIT compliance following the current audit? No evidence, to the best of our knowledge, on this to date exists. It would be therefore desirable to exploit the information from other CIT waves with a view of disentangling the impact of having been audited again in the past on present and future tax compliance. This is by no means an easy issue to explore, but it is certainly one that deserves attention.

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

[64] As already mentioned, in order to provide additional evidence and support the external validity of the analysis, Appendix E.1 performs the estimation of both the aggregate and the type-specific $ATT$ for the 2013 wave, generally corroborating the results presented in Section 6.

# 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, Appendices B and C present the results of several additional sensitivity analyses performed testing the robustness of the findings as already described in Section 5.1 (see in particular footnotes 53, 51 and 57). In this light, we follow two main avenues. First, we extend the methodologies discussed in the main text by incorporating regression specifications for the outcome variables (Appendix B). Secondly, we perform a stricter selection of the matched sample through the CEM methodology by employing a less parsimonious set of covariates for stratification and subsequently implement our baseline models (Appendix C). Additionally, Appendix D replicates the estimation of the audit-type specific ATT (Section 6.2) separately for two clusters of businesses with different probability of being audited. Finally, Appendix E presents further validations of our estimation strategy. This is achieved through a) replicating the main analysis for the 2013 wave (Subsection E.1), b) testing the sensitivity of results reported in Section 6 to random subsampling and subsampling targeting outliers (Subsection E.2), and (c) by providing a placebo impact estimate for a cluster of taxpayers audited after their 2016 tax return was reported treating them as they were audited in the 2015 wave (Subsection E.3).

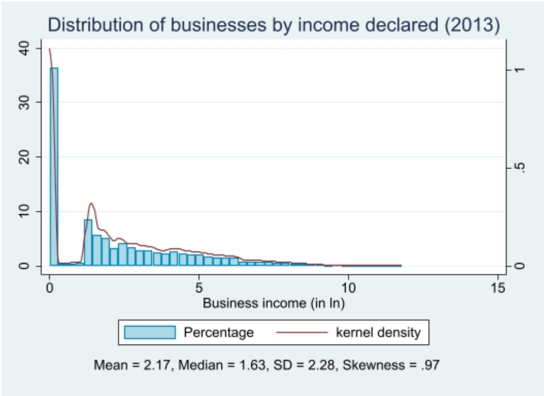
# Appendix A: Additional figures

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.

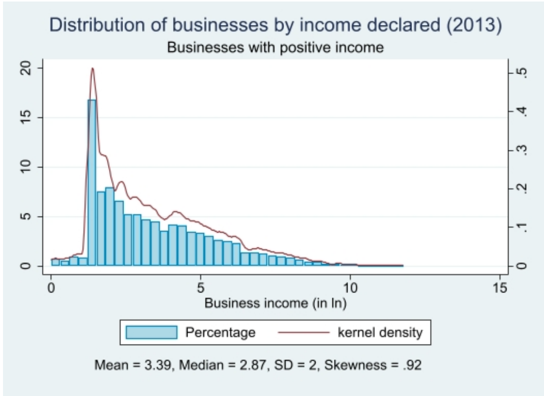
Figure A.2 reports the distribution of taxable income by deciles of population for the universe of CIT filers and by size of firms for the same selected years as in Figure A.1. Note that in terms of CTI reported, firms in the tenth decile declare more than 90% of taxable income across all periods in 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% of the total audits versus 41% of comprehensive and in 2015 narrow-scope audits are 63% of the total versus 37% of comprehensive. The 2013 wave is a bit more of an outlier with a 52% of narrow-scope audits and a 48% 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% of the total) following with audits on Large-businesses (19.65%).

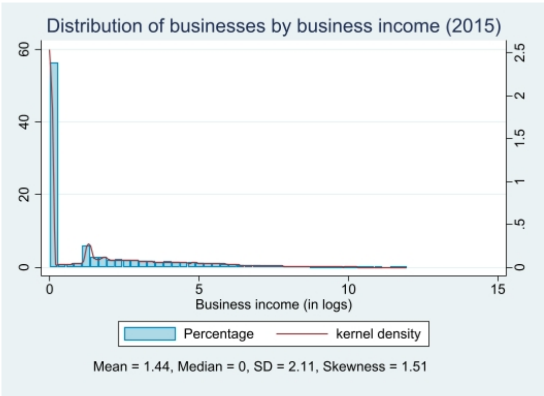
Figure A.1: Distribution of businesses by business income declared (for years 2013, 2015 and 2018)



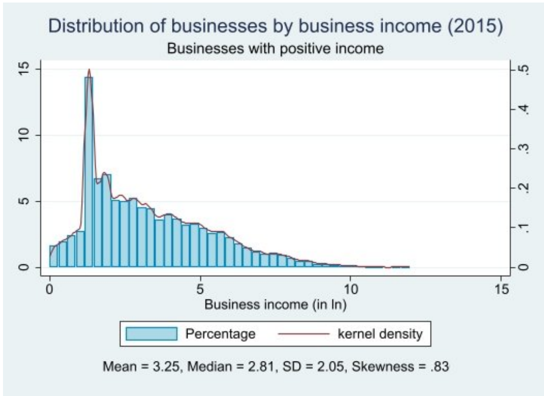

(a)



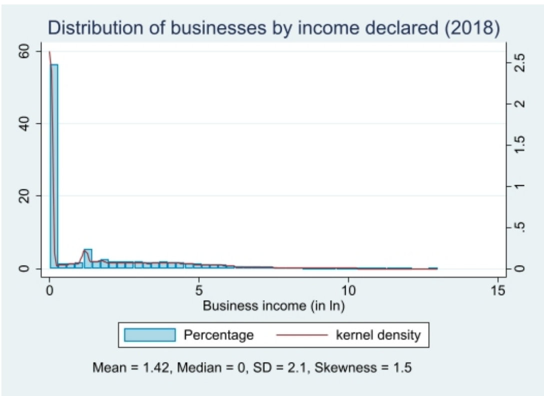

(b)



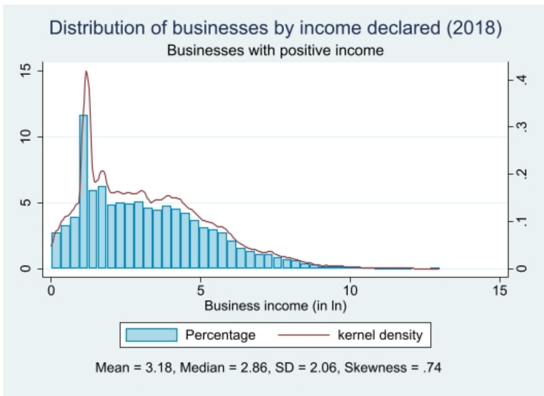

(c)




(d)




(e)

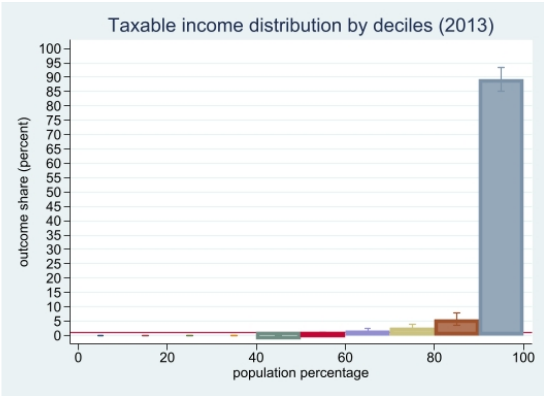



(f)

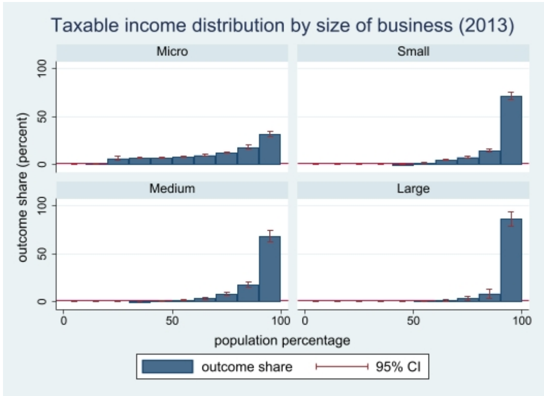
Note: Authors' calculations based on data provided by RRA.

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, businesses in the last decile of the CTI and taxpayers not reporting positive tax liabilities (including nil-filers and firms reporting losses), with small businesses reporting positive

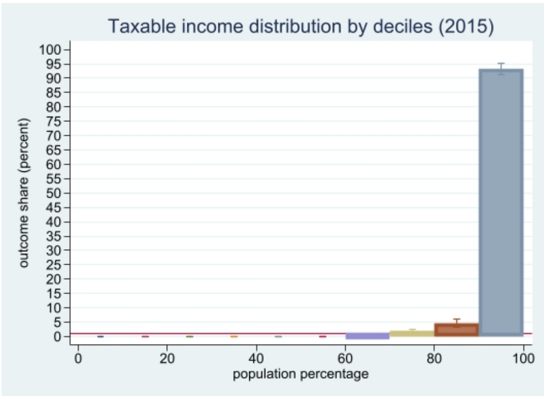
Figure A.2: Distribution of CTI reported (for years 2013, 2015, and 2018)



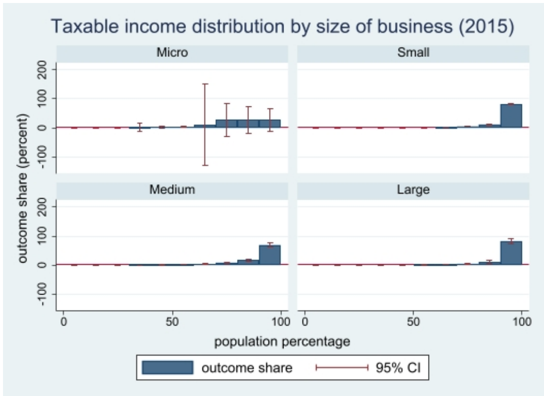

(a)



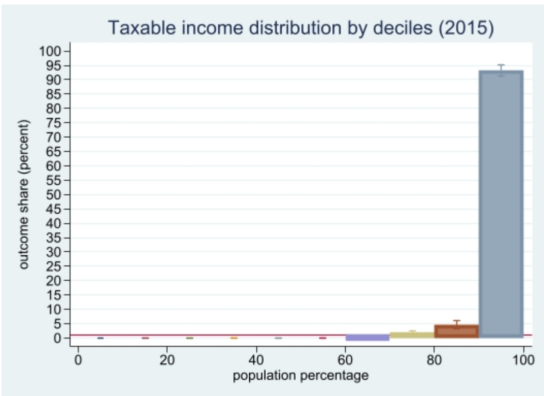

(b)



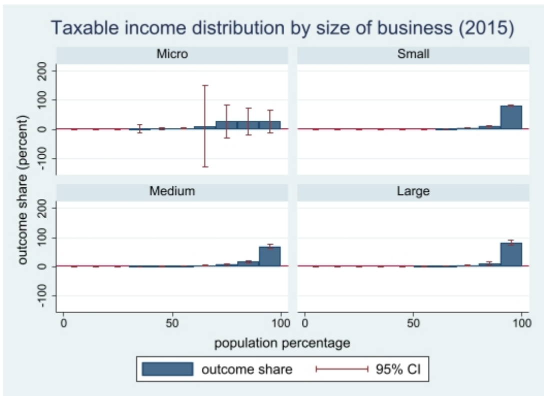

(c)




(d)




(e)



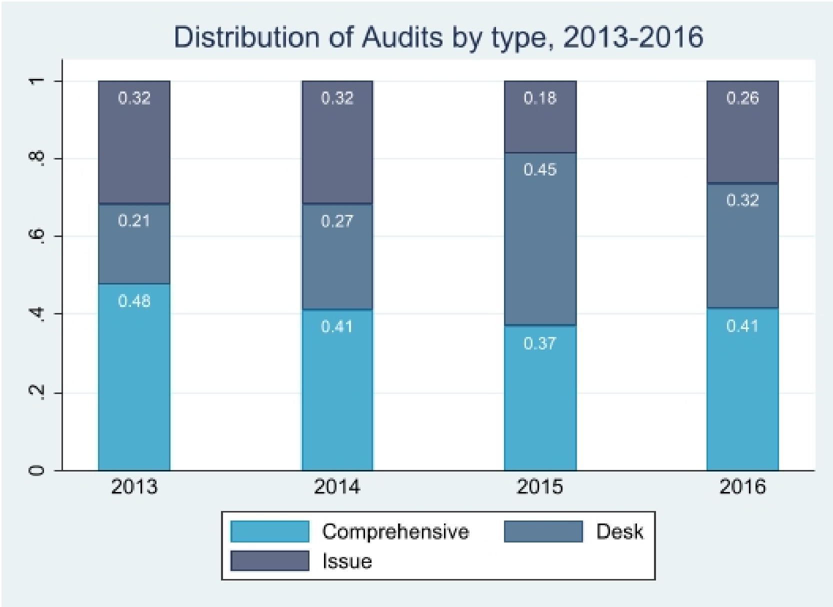

(f)

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

CTI generally less likely to be audited. Thus, CTI seems to play a role in audit selection but in a non-linear fashion since the probability of being audited and the CTI present a U-shaped relationship. Now we explore the relationship between the probability of being audited and the risk scores, the other crucial dimension of our matching strategy.

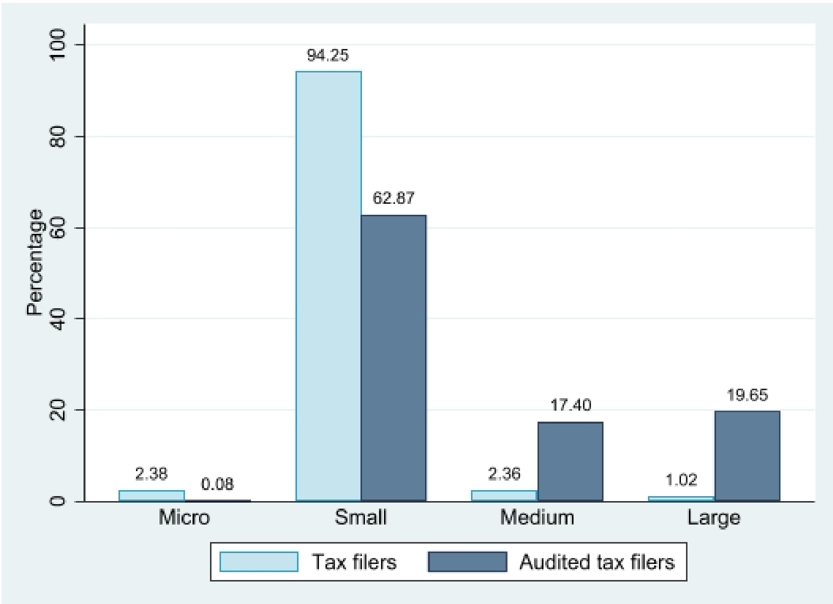
Figure A.6 plots the probability of being audited across deciles of the 2015 Risk Index which is increasing. The correlation between the deciles of the 2015 Risk Index and

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




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

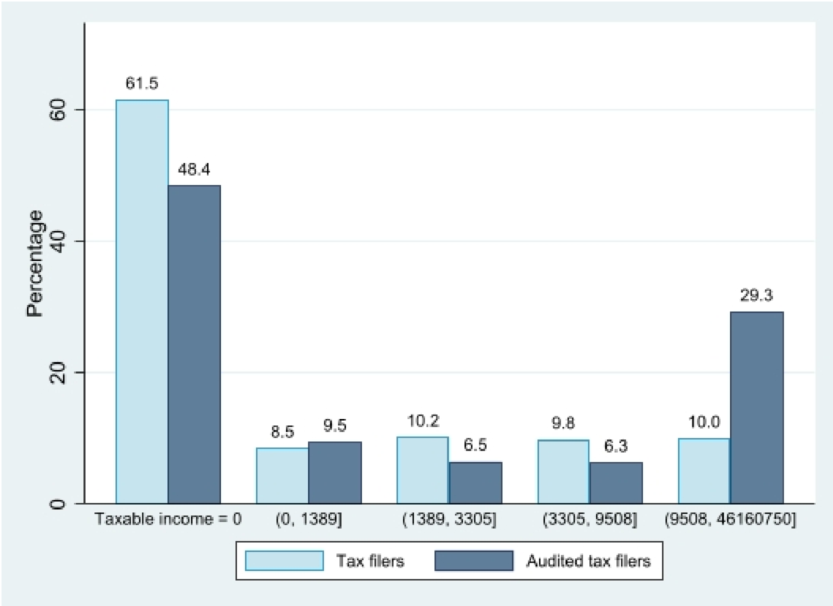
Figure A.4: Distribution of businesses and audits by size (Average 2013-2016), %




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

the probability of being audited within each decile in the 2015 audit wave is calculated to be 0.9113. What all this suggests is that the Risk Index plays an important role in the audit selection process. As already mentioned in the main text (see footnote 50), 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

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

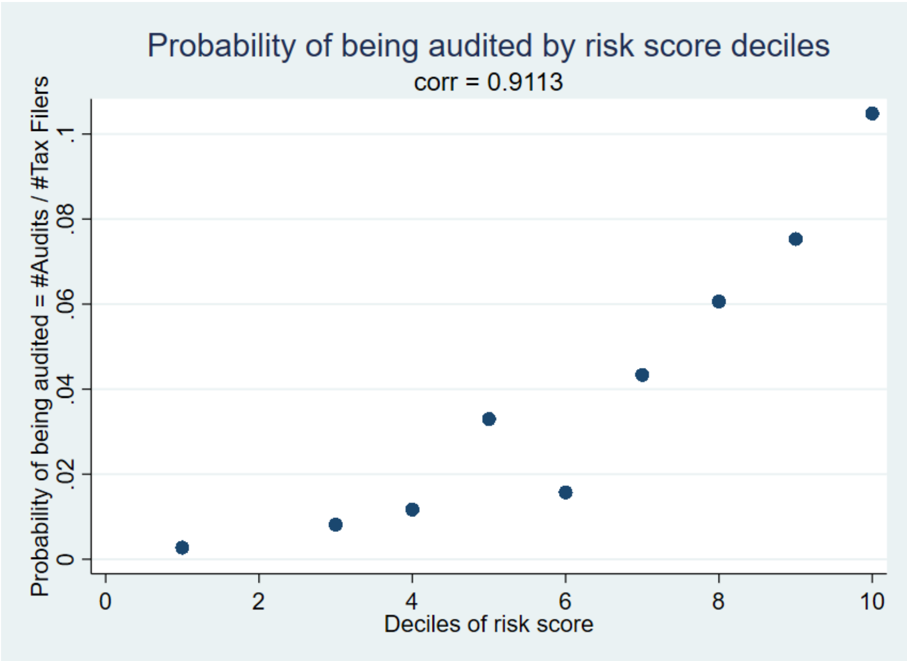



Note: Authors' calculations based on data provided by RRA.

in the control group). This explains the lack of a clear cutoff point in the likelihood Risk Index which determines the set of businesses to be audited.[65] 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 CTI reported. This figure suggests that the probability of being audited correlates with the combination of our two main matching variables in a non-linear way confirming the role played by CTI (Figure A.5). Indeed, peaks of auditing probabilities can be generally found at the highest deciles of risk scores both for top income declarers and businesses reporting nil tax liabilities 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. Also, this seems to indicate that CTI reported might pick up a significant part of the variability in audit probability due to different likelihood of audits to lead to the expected revenue target.

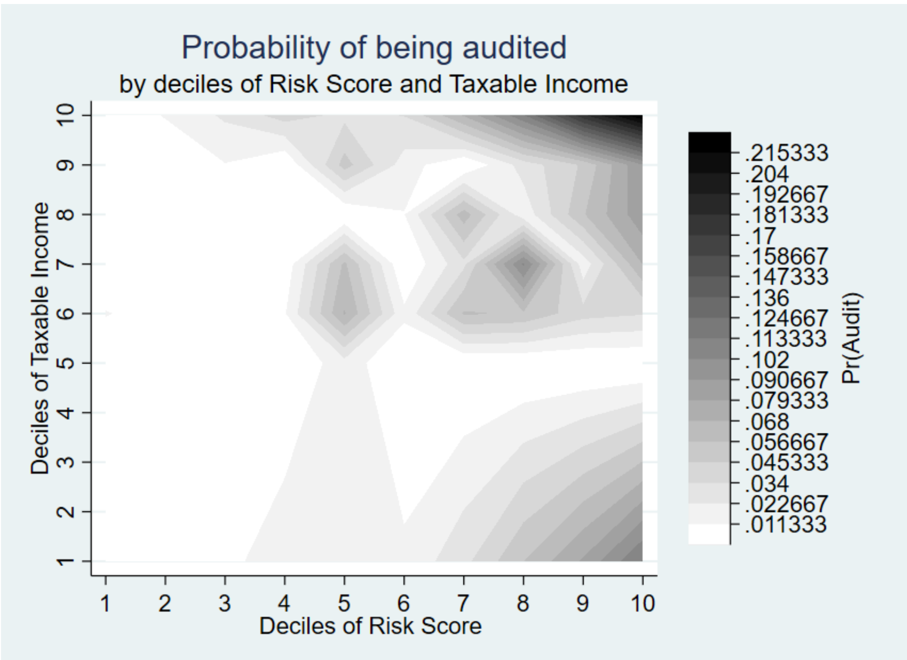
[65]For this reason we cannot employ an RDD.

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 CTI reported




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

# Appendix B: Regression adjustment models

The use of approximate matching techniques might imply some residual imbalance in the matched data.[66] Thus, a reasonable approach to deal with this is to adjust for the potential remaining imbalance in the estimation of the aggregate $ATT$ via a regression model. We do so by employing both weighted regression models based on the weights calculated with our baseline models (Table B.1) and double-robust regression adjustment models (Table B.2) based on the inverse probability of treatment (IPW-RA) simultaneously calculated with the outcome model via a logit model. This latter estimation technique is double-robust 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).

Table B.1: Main Results – Aggregate $ATT$ (weighted regression models)

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Matching estimator | (1) | (2) | (3) | (4) | (5) | (6) |
| CEM | 0.321*** | 0.243 | 0.234 | 0.214** | 0.228 | 0.173 |
| | (0.095) | (0.188) | (0.202) | (0.090) | (0.142) | (0.161) |
| Kernel - MHD | 0.269*** | 0.007 | 0.029 | 0.166*** | 0.045 | 0.018 |
| | (0.100) | (0.135) | (0.114) | (0.061) | (0.083) | (0.082) |
| Kernel - PSM | 0.187* | -0.038 | -0.073 | 0.133** | 0.057 | -0.011 |
| | (0.095) | (0.125) | (0.132) | (0.063) | (0.086) | (0.091) |
| Nearest Neighbour | 0.627*** | 0.324 | 0.265 | 0.383** | 0.207 | 0.172 |
| | (0.222) | (0.307) | (0.276) | (0.182) | (0.234) | (0.170) |

Note: Note: Standard errors are reported in parentheses. In particular, robust standard errors (clustered by tax center) are reported for CEM estimator, bootstrapped standard errors based on 200 replications are reported for Kernel - MHD and Kernel - PSM estimators, and wild bootstrapped standard errors based on 200 replications are reported for Nearest Neighbour estimator; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls are decribed in Appendix B while matching variables are described in Section 5.1.

For the weighted regression outcome models we employ as covariates a dummy indicating whether the taxpayer adopted the EBM before the 2015 audit wave—which represents a potential confounding event in terms of future deterrence of audits (we return on this also in Appendix C)—the risk score assigned to the taxpayer each of the three years before treatment, the taxable income reported in 2014 and 2013, the VAT paid on inputs reported each of the three years before treatment, a set of indicator variables for the tax centre, the sector of activity and the finer classification of the section of activity (according to the ISIC classification), dummies for diverse type of income reported each of the three years before treatment and a dummy for CIT return reported after the deadline during the year of the audit process. For the regression adjustment models we

[66]For example, by using CEM the remaining imbalance on the matching covariates depends on the remaining variation within the coarsened bins.

employ two alternative set of covariates both including a dummy indicating whether the taxpayer adopted the EBM before the 2015 audit wave, the risk scores for the latest two pre-treatment years, reported taxable income declared in the year before treatment and a dummy for the sector of activity. The more inclusive set of covariates (Set II) also includes dummies for diverse type of income reported each of the three years before treatment, a dummy for CIT tax return reported after the deadline during the year of the audit process and a dummy identifying the three tax centers in Kigali.[67] These analyses corroborate our main findings: by controlling for comprehensive sets of covariates affecting taxpayers' reporting behaviour and more generally firms' business cycle, the results remain qualitatively unchanged and quantitatively coherent with the ones presented in the main text.

Table B.2: Main Results – Aggregate ATT (regression adjustment models)

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Matching estimator | (1) | (2) | (3) | (4) | (5) | (6) |
| IPW-RA (set I) | 0.187** | 0.072 | 0.080 | 0.137* | 0.115 | 0.087 |
| | (0.093) | (0.182) | (0.127) | (0.076) | (0.138) | (0.098) |
| IPW-RA (set II) | 0.168* | 0.082 | 0.076 | 0.127** | 0.125 | 0.091 |
| | (0.060) | (0.166) | (0.096) | (0.053) | (0.126) | (0.077) |

Note: Robust standard errors (clustered by tax center) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls are described in Appendix B while matching variables are described in Section 5.1.

In the same light, to estimate the $ATT$ by audit type, we rely on double-robust regression adjustment versions of our baseline IPTW model. For the outcome model we employ two alternative sets of covariates. Set I includes a dummy indicating whether the taxpayer adopted the EBM before the 2015 audit wave, taxable income reported in 2014, the risk score for the last two pre-treatment years and a dummy for the sector of activity (see Table B.3). Set II also includes dummies for diverse type of income reported each of the three years before treatment, a dummy for CIT return reported after the deadline during the year of the audit process and a dummy identifying the four tax centers in Kigali (see Table B.4).[68] As for the other regression adjustment models, IPTW-RA estimators use a model to predict treatment status, and they use another model to predict outcomes. As already mentioned, since these regression adjusted estimators have the double-robust property, only one of the two models must be correctly specified for the IPWRA estimator to be consistent and thus there is no need to correct the standard errors in the second step to reflect the uncertainty surrounding the predicted outcomes (Cattaneo, 2010; Wooldridge, 2002, 2007). Also in this case, the results of the analysis are not sensitive to changes due to potential confounding factors and validate our main findings both qualitatively and quantitatively.

[67] The sets of covariates are less inclusive with respect to the weighted regression models in order to ensure achieving the convergence of the model.

[68] The selection of these variables is based on their predictive power while ensuring the achievement of the convergence of the model.

Table B.3: Main Results – *ATT* by audit type (IPTW-RA), Set I

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Type of Audit | (1) | (2) | (3) | (4) | (5) | (6) |
| Comprehensive | 0.371* | 0.190 | 0.126 | 0.305** | 0.184 | 0.165 |
| | (0.197) | (0.223) | (0.290) | (0.150) | (0.186) | (0.214) |
| Narrow-scope | 0.019 | -0.238*** | -0.175*** | 0.006 | -0.097** | -0.082** |
| | (0.031) | (0.066) | (0.046) | (0.027) | (0.047) | (0.041) |

Note: Robust standard errors (clustered by tax center) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls are described in Appendix B while matching variables are described in Section 5.1.

Table B.4: Main Results – *ATT* by audit type (IPTW-RA), Set II

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Type of Audit | (1) | (2) | (3) | (4) | (5) | (6) |
| Comprehensive | 0.276** | 0.142 | 0.134 | 0.230** | 0.144 | 0.176 |
| | (0.140) | (0.167) | (0.281) | (0.105) | (0.138) | (0.207) |
| Narrow-scope | 0.018 | -0.231*** | -0.168*** | 0.007 | -0.092* | -0.076* |
| | (0.031) | (0.069) | (0.042) | (0.028) | (0.050) | (0.040) |

Note: Robust standard errors (clustered by tax center) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls are described in Appendix B while matching variables are described in Section 5.1.

# Appendix C: 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 5.1. Here we present the results of two of them. Tables C.1 and C.2 present the results of the analysis based on Set II of matching covariates which includes the initial set of control variables described in Section 5.1 and add the amount of losses reported by the business and an indicator variable for a business's adoption of EBM before the start of the 2015 audit wave (exact matching). In particular, the adoption of EBM by the businesses has been not uniform in time across firms and appears to be a significant improvement in RRA enforcement capacity. Thus exposure to EBMs could be a potential confounder for the impact of audits on future reporting behaviour. Appendix B already accounted for this issue indirectly by using this variable as a control in the outcome models finding no significant alteration in the results. Here, by exact matching on the timing of adoption of EBMs across treatment cohorts we directly ensure the internal validity of the estimates by explicitly providing that exposure to EBMs does not differ across the comparison groups in the matched sample.

The matched set of observations includes 280 treated units (77.4% of the pre-treatment units) and 5072 untreated units (46.7%), 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 62.9% of the initial imbalance because it is inherently more difficult to reduce imbalance simultaneously on an extended set of variables. Nevertheless, the univariate imbalance metrics reduce substantially.[69] The results of the analysis corroborate our main findings both qualitatively and quantitatively. In particular, the average effect on the aggregate across methods is 19% (11.9%) increase on CTI (CIT) reporting by audited taxpayers one year after the audit process and compared to the matched control group (see Table C.1). Also regarding the type-specific ATT the analysis confirms the results both in terms of the sign of the estimated effects and their magnitude (see Table C.2). In particular, in this case, while the negative impact of Narrow-scope audits on reporting behaviour persists 2 and 3 years after the process, the relatively low positive impact of Narrow-scope audits recorded for the first year after the process (3.6%) turns out to be statistically different from zero at least for CTI reported.

By using a larger set of stratification covariates that includes set II of covariates and adds VAT paid on inputs declared each of the two years before treatment, an indicator variable for tax centres in Kigali (exact matching) and a dummy for the broad sector of activity (exact matching), the matched set of observations include 201 audited taxpayers (56% of the pre-treatment units) and 3105 untreated units (29% of the pre-treatment units). Thus, by increasing even more the stratification criteria the sample size reduces drastically. Multivariate imbalance measure on this whole set of covariates after CEM reduces to 86% of the initial level while univariate imbalance on the variables considered

[69] In particular, imbalance on the adoption of EBM is completely eliminated by exact matching, while imbalance on the Risk Index reduces to 20.5% of the initial level. Regarding the rest of variables, imbalance reduces to 50%, 34%, and 43% for taxable income reported in 2013, 2014 and 2015 respectively and to 70% for the losses reported which had an initial level of imbalance already relatively low (0.12).

Table C.1: Main Results – Aggregate $ATT$ (using Set II of matching covariates)

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Matching estimator | (1) | (2) | (3) | (4) | (5) | (6) |
| CEM | 0.142*** | -0.019 | -0.043 | 0.089*** | 0.025 | -0.035 |
| | (0.039) | (0.123) | (0.105) | (0.016) | (0.088) | (0.074) |
| Kernel - MHD | 0.195*** | -0.008 | -0.011 | 0.104* | 0.015 | -0.014 |
| | (0.075) | (0.109) | (0.103) | (0.056) | (0.081) | (0.071) |
| Kernel - PSM | 0.178* | -0.066 | -0.095 | 0.133* | 0.048 | -0.007 |
| | (0.108) | (0.139) | (0.133) | (0.080) | (0.097) | (0.094) |
| Nearest Neighbour | 0.244** | 0.091 | 0.021 | 0.148** | (0.092) | -0.003 |
| | (0.095) | (0.165) | (0.169) | (0.066) | (0.120) | (0.122) |

Note: Standard errors are reported in parentheses. In particular, robust standard errors (clustered by tax center) are reported for CEM estimator, bootstrapped standard errors based on 200 replications are reported for Kernel - MHD and Kernel - PSM estimators and heteroskedasticity-consistent analytical standard errors proposed by Abadie and Imbens (2006, 2008) are reported for Nearest Neighbour estimator; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Set II of matching covariates includes the initial set of control variables described in Section 5.1 and dummies for the sector of activity (according to ISIC classification). The matched set of observations include 280treated units (76.4%) and 5072 untreated units (46.7%).

Table C.2: Main Results – $ATT$ by audit type (using Set II of matching covariates)

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Type of Audit | (1) | (2) | (3) | (4) | (5) | (6) |
| Comprehensive | 0.235** | -0.034 | -0.160 | 0.198*** | 0.012 | -0.022 |
| | (0.104) | (0.225) | (0.246) | (0.073) | (0.174) | (0.163) |
| Narrow-scope | 0.036* | -0.262*** | -0.201*** | 0.017 | -0.112** | -0.099** |
| | (0.022) | (0.064) | (0.046) | (0.020) | (0.047) | (0.040) |

Note: Robust standard errors (clustered by tax center) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Set II of matching covariates includes the initial set of control variables described in Section 5.1 and dummies for the sector of activity (according to ISIC classification). The matched set of observations include 280treated units (76.4%) and 5072 untreated units (46.7%).

in isolation reduces substantially.[70] All this indicates that it is more difficult to find control units in the stratification process when using an extended set of variables in particular when exact matching on a subset of them. Nevertheless, Table C.3 and C.4

[70]Namely, imbalance on the Risk Index reduces to 31.5% of initial level, imbalance on the variables for which exact matching is performed (indicators for the adoption of EBM, the ISIC sector and the tax centers in Kigali) is completely eliminated and imbalance reduces to 35.7%, 30.3%, and 33.5% for taxable income reported in 2013, 2014 and 2015 respectively and to 64.6% for the losses reported which had an initial level of imbalance already relatively low (0.13). Finally imbalance reduces to 52% and 56% for VAT paid on inputs in 2014 and 2015 respectively.

present the results of this analysis that are again consistent with our main findings. More precisely, when focusing on these subset of audited taxpayers for which it is possible to find matches on more dimensions than those provided in the main text or by using set II of covariates, we find qualitatively the same results and quantitatively generally a slightly higher but comparable magnitudes of the estimated effect on the aggregate[71] and a comparable magnitudes for the type-specific impact of audits which also in this case presents a negative impact the second and third year after the enforcement occurred while showing a limited positive but significant impact on CTI reported the first year after the process. This whole robustness analysis shows that performing the CEM stratification on a larger set of covariates reduces (sometimes drastically) 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.

Table C.3: Main Results – Aggregate $ATT$ (using Set III of matching covariates)

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Matching estimator | (1) | (2) | (3) | (4) | (5) | (6) |
| CEM | 0.342*** | 0.259** | 0.253** | 0.208** | 0.191* | 0.143 |
| | (0.121) | (0.119) | (0.104) | (0.091) | (0.097) | (0.096) |
| Kernel - MHD | 0.178** | 0.072 | 0.035 | 0.117* | 0.071 | 0.011 |
| | (0.090) | (0.130) | (0.115) | (0.061) | (0.089) | (0.079) |
| Kernel - PSM | 0.222** | -0.022 | -0.075 | 0.174** | 0.084 | 0.020 |
| | (0.103) | (0.141) | (0.155) | (0.071) | (0.098) | (0.105) |
| Nearest Neighbour | 0.329** | 0.026 | 0.246 | 0.239** | 0.069 | 0.108 |
| | (0.146) | (0.174) | (0.196) | (0.104) | (0.118) | (0.145) |

Note: Standard errors are reported in parentheses. In particular, robust standard errors (clustered by tax center) are reported for CEM estimator, bootstrapped standard errors based on 200 replications are reported for Kernel - MHD and Kernel - PSM estimators and heteroskedasticity-consistent analytical standard errors proposed by Abadie and Imbens (2006, 2008) are reported for Nearest Neighbour estimator; $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$. Set III of matching covariates includes set II covariates and VAT paid on inputs declared each of the two years before treatment and dummies for the tax centre. The matched set of observations include 201 treated units (56%) and 3105 untreated units (29%).

[71] The average effect of audits on the aggregate one year after the process and across methods corresponds to an increase of 26.8% (18.5%) in reported CTI (CIT).

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

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Type of Audit | (1) | (2) | (3) | (4) | (5) | (6) |
| Comprehensive | 0.315* | 0.067 | -0.213 | 0.279** | 0.098 | -0.047 |
| | (0.176) | (0.239) | (0.293) | (0.126) | (0.156) | (0.151) |
| Narrow-scope | 0.048*** | -0.187*** | -0.212*** | 0.021 | -0.067* | -0.127*** |
| | (0.016) | (0.054) | (0.033) | (0.021) | (0.038) | (0.027) |

Note: Robust standard errors (clustered by tax center) are reported in parentheses; $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$. Set III of matching covariates includes set II covariates and VAT paid on inputs declared each of the two years before treatment and dummies for the tax centre. The matched set of observations include 201 treated units (56%) and 3105 untreated units (29%).

# Appendix D: Audit-type specific *ATT* by groups of businesses

This appendix replicates the analysis presented in Section 6.2 by clustering businesses in two groups and distinguished by their probability of being audited. Group 1 consists of taxpayers reporting null tax liabilities (Nil-filers and firms reporting losses and thus null taxable income) across all sizes and *Medium-Large* businesses declaring positive income. This whole cluster groups about 65.5% of audited businesses in our matched sample (13.8% are *Medium-Large* businesses and 57.5% are businesses reporting null tax liabilities). Businesses with high risk scores in this group face a relatively high probability of being audited (see Figure A.7). It is reasonable to assume that large businesses with high risk score for the likelihood of underreporting also have high score for the potential impact of uncovered tax noncompliance. The same reasoning applies to businesses reporting null tax liabilities: both nil-filing and reporting losses may be associated behaviour consistent with substantial underreporting. Group 2 consists of the residual set of *Small* businesses declaring positive income (34.5% of audited businesses) generally facing lower probabilities of being audited also at higher deciles of risk scores (see Figure A.7). These businesses are likely to be associated with a low score for the potential impact on revenues. Tables D.1 and D.2 present the results of the estimation of the *ATT* of audits by audit type for businesses in group 1 and 2 respectively. These results seem to suggest that both audit strategies have different impact depending on the type of businesses audited. Indeed, while both comprehensive and narrow-scope audits have a pro-deterrence effect on the behaviour of businesses in the first group (Table D.1) the first year after the enforcement[72], they have both a counter-deterrence effect on businesses belonging to the second group (Table D.2). Moreover, for these businesses the negative impact of comprehensive audits is persistent when time passes. This analysis suggests that the performance of audits is not uniform across types of businesses and that given the different probabilities of being audited firms face across groups, mechanisms related to the role of information provided by audits as described in Section 3 might play a part in determining this results.

Table D.1: Main Results – *ATT* by audit type (IPTW), Group 1: Nil-filers (all sizes) & Medium-Large firms declaring positive income

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Type of Audit | (1) | (2) | (3) | (4) | (5) | (6) |
| Comprehensive | 0.267** | 0.072 | -0.105 | 0.252*** | 0.117 | -0.020 |
| | (0.111) | (0.226) | (0.165) | (0.092) | (0.187) | (0.122) |
| Narrow-scope | 0.199*** | 0.005 | 0.115 | 0.081*** | -0.014 | 0.023 |
| | (0.061) | (0.103) | (0.073) | (0.026) | (0.085) | (0.076) |

Note: Robust standard errors (clustered by tax center) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Matching variables are described in Section 5.1.

[72] The same results hold also separately for the 2 subgroups of which is composed this group implicitly providing a robustness to outliers.

Table D.2: Main Results – $ATT$ by audit type (IPTW), Group 2: Small firms declaring positive income

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Type of Audit | (1) | (2) | (3) | (4) | (5) | (6) |
| Comprehensive | -0.524*** | -0.822*** | -1.459*** | -0.396*** | -0.516*** | -0.158*** |
| | 0.132) | (0.042) | (0.257) | (0.092) | (0.025) | (0.056) |
| Narrow-scope | -0.052 | -0.322*** | -0.125 | -0.035 | -0.130** | -0.019 |
| | 0.107) | (0.070) | (0.085) | (0.083) | (0.063) | (0.076) |

Note: Robust standard errors (clustered by tax center) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Matching variables are described in Section 5.1.

# Appendix E: Further robustness analysis

## E.1: Audit wave 2013

With the aim of checking whether the results obtained for audit wave 2015 are generalisable to other audit waves, this appendix replicates the analysis by estimating the $ATT$ of audit wave 2013 on CTI and CIT reporting behaviour. This wave comprises of a significantly lower number of audits performed compared to the 2015 wave (257 vs. 435). Applying all the steps of the sample selection performed in the main analysis gives a matched sample of significantly reduced dimensions due to the non-audit restrictions in following waves.[73] For this reason, we do not apply those specific restrictions here and we perform the CEM initial stratification just on businesses' Risk Score and tax liabilities reported the year before the treatment is applied. By doing so we are able to keep 205 audited taxpayers in the matched sample. A direct consequence of applying the relaxed restrictions is that the impact estimated 2 and 3 years after the audit process is likely to be affected by the confounding effects embedded in the following audit waves. Table E.1 presents the results on the aggregate $ATT$: across matching methods the aggregate impact on CTI (CIT) reported the first year after the audit process is 23.6% (18.5%) which is qualitatively and quantitatively comparable to the results obtained for the 2015 wave. As expected, the impact estimated for the following 2 years is significant, higher in magnitude, and increasing across time as a result of the combined effects of different audit processes. As already noted in the concluding remarks, trying to disentangle the impact of audits targeting the same population more than one year is by no means an easy task and this is beyond the scope of this paper. However, the evidence emerging suggests that repeated audit targeting has a positive impact on deterrence going beyond the first year after the enforcement process started.

Table E.2 shows the results of the analysis performed by type of audits which corroborates the results of the analysis implemented on 2015 wave. In particular, the impact of comprehensive audits seems to confirm the reasoning outlined above: there is a significant pro-deterrence effect of this type of audits—driving the results on the aggregate—that goes beyond the first year and increases over time as a result of the combining effects from different waves of audits. As in the main analysis, narrow-scope audits have a counter-deterrence effect on audited taxpayers' future reporting behaviour. For the case of 2013 audit wave the coefficients are always negative and they are significant for the first and, at least for CTI, the second year after the audit.

[73] As already discussed, the quality of data for the 2015 wave was the main criterion used to select that wave for the main analysis.

Table E.1: Main Results – Aggregate $ATT$ (2013 audit wave)

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Matching estimator | (1) | (2) | (3) | (4) | (5) | (6) |
| CEM | 0.217* | 0.427** | 0.438** | 0.162* | 0.318** | 0.336** |
| | (0.126) | (0.171) | (0.196) | (0.090) | (0.126) | (0.144) |
| Kernel - MHD | 0.225* | 0.437*** | 0.478** | 0.176* | 0.293** | 0.316** |
| | (0.126) | (0.156) | (0.204) | (0.093) | (0.116) | (0.152) |
| Kernel - PSM | 0.244* | 0.397** | 0.310 | 0.196** | 0.300** | 0.244* |
| | (0.125) | (0.168) | (0.199) | (0.095) | (0.129) | (0.147) |
| Nearest Neighbour | 0.258** | 0.321** | 0.387** | 0.207** | 0.277** | 0.319** |
| | (0.123) | (0.155) | (0.178) | (0.090) | (0.116) | (0.137) |

Note: Standard errors are reported in parentheses. In particular, robust standard errors (clustered by tax center) are reported for CEM estimator, bootstrapped standard errors based on 200 replications are reported for Kernel - MHD and Kernel - PSM estimators and heteroskedasticity-consistent analytical standard errors proposed by Abadie and Imbens (2006, 2008) are reported for Nearest Neighbour estimator; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E.2: Main Results – $ATT$ by audit type (2013 audit wave)

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Type of Audit | (1) | (2) | (3) | (4) | (5) | (6) |
| Comprehensive | 0.489** | 0.715*** | 0.894** | 0.405** | 0.574*** | 0.763** |
| | (0.198) | (0.213) | (0.363) | (0.162) | (0.153) | (0.338) |
| Narrow-scope | -0.184*** | -0.099* | -0.081 | -0.102** | -0.081 | -0.064 |
| | (0.046) | (0.057) | (0.113) | (0.051) | (0.074) | (0.099) |

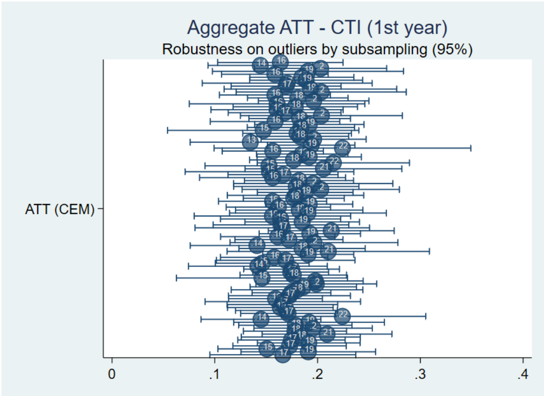
Note: Robust standard errors (clustered by tax center) are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## E.2: Robustness on outliers

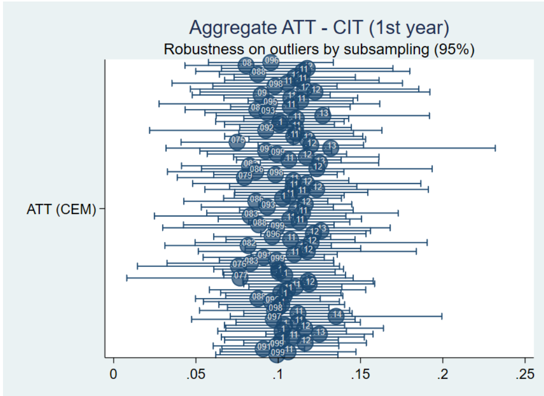
In order to test how sensitive are our results to outliers, we draw 100 subsamples out of our matched dataset in a way that each subsample keeps a different randomly selected 95% of the treated sample and all their matched controls and we perform the estimation of the $ATT$ on those samples. Figure E.2 plots the results of these subsampling analysis for the aggregate $ATT$ each of the three years after the audit process. For the first year of impact, the coefficient estimates for CTI (CIT) span from an impact of 13% (7.5%) to an impact of 22% (14%) but tend to concentrate in the middle point of this interval, close to the correspondent estimate for the whole sample (see Table 8, Columns 1 and 4) and they are always significant at least at the 5% level. The second and third year the estimated impact tend to concentrate closer to zero and they are never significant as happens on the whole sample (see Table 8, Columns 2, 3, 5, and 6). Figure E.2 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 narrow-scope audits is very close to and always statistically indistinguishable from zero while the impact of comprehensive audits tend to concentrate close to the effect estimated for the whole sample and it is in the large majority of cases statistically significant at least at the 10% level. The second and third year after the enforcement started, the impact of comprehensive audits become statistically indistinguishable from zero 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 with very low variability and a level of significance of 5% at least. Overall, this random subsampling exercise seems to corroborate that our main results are not sensitive to outliers.

Nevertheless, in order to further confirm this and rule out the possibility that our results are driven by very large firms we perform a trimming exercise by dropping from the matched observations the top 5% of the treated sample in terms of their declared CTI and all their matched controls. Table E.3 reports the estimates of this exercise for the aggregate $ATT$ while Table E.4 those for the audit-specific impact. In both cases the results are qualitatively and quantitatively comparable to the ones presented in Section 6 corroborating that our analysis is not sensitive to outliers in terms of their turnover and that the results are not driven by very large firms.

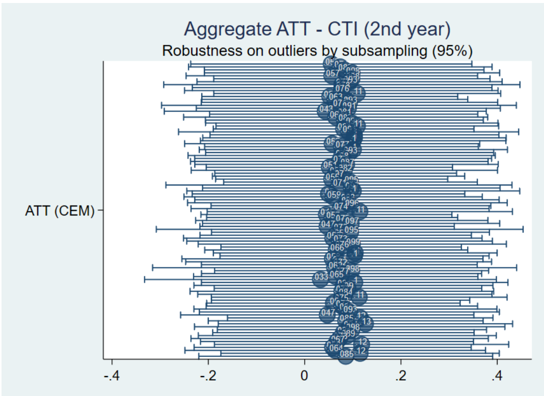
Figure E.2: Random subsampling on 95% of treated (and their matches)



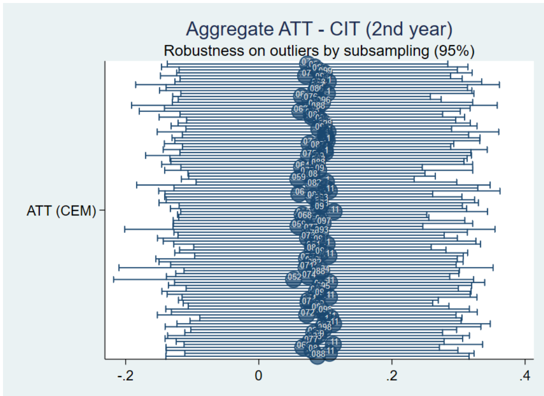

(a)



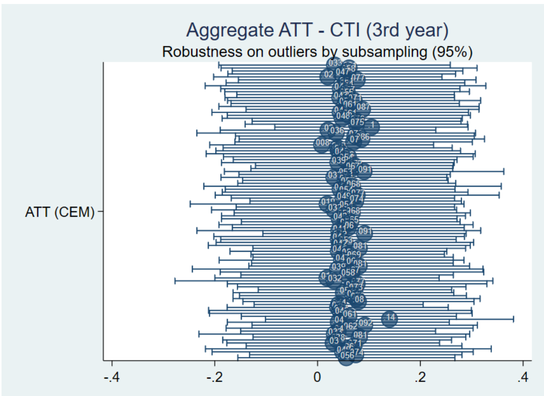

(b)



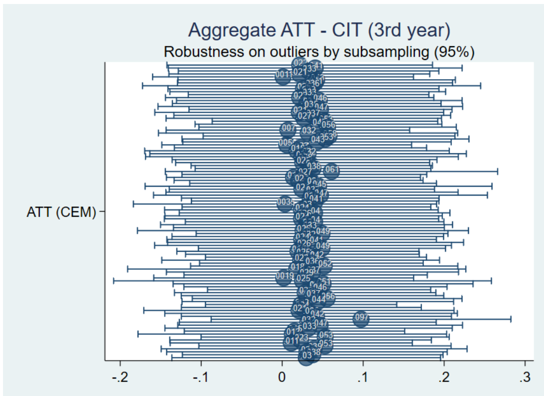

(c)




(d)



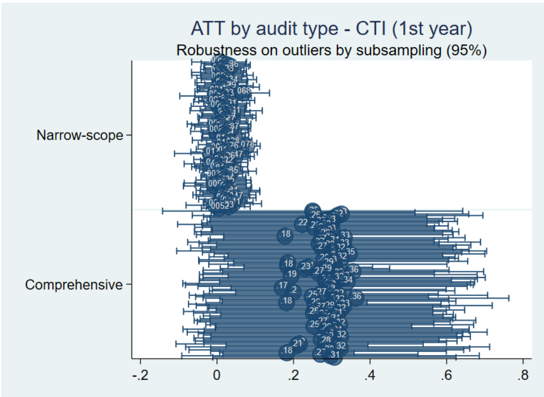

(e)



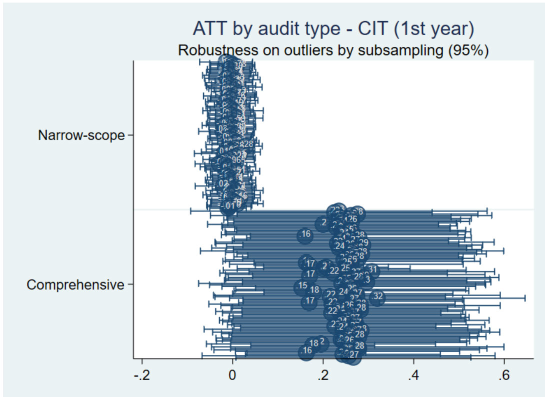

(f)

Note: These graphs represent, by outcome variable (CTI and CIT) and year of the impact (I, II, and III), the results of the estimation of the aggregate $ATT$ obtained using CEM in 100 random subsamples including 95% of the treated taxpayers and their matched control units. Estimates are depicted with 95% confidence intervals.

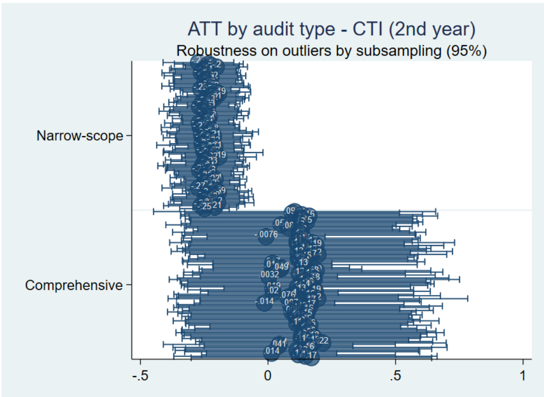
Figure E.2: Random subsampling on 95% of treated (and their matches)



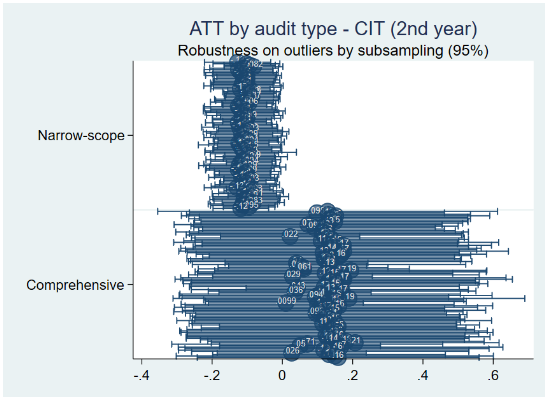

(a)



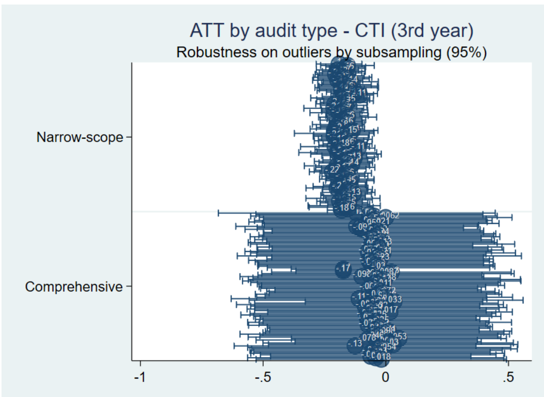

(b)



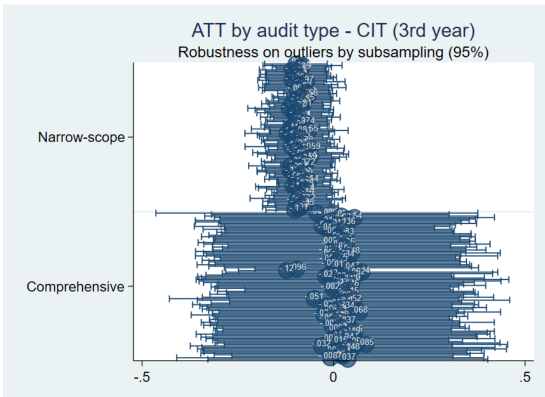

(c)




(d)




(e)




(f)

Note: These graphs represent, by outcome variable (CTI and CIT) 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 sumsamples including 95% of the treated taxpayers and their matched control units. Estimates are depicted with 95% confidence intervals.

Table E.3: Main Results – Aggregate $ATT$ (trimming top 95% of treated and matched controls)

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Matching estimator | (1) | (2) | (3) | (4) | (5) | (6) |
| CEM | 0.228*** | 0.054 | 0.046 | 0.145** | 0.069 | 0.024 |
| | (0.071) | (0.154) | (0.136) | (0.058) | (0.110) | (0.098) |
| Kernel - MHD | 0.228*** | 0.090 | 0.097 | 0.116** | 0.086 | 0.066 |
| | (0.082) | (0.098) | (0.100) | (0.057) | (0.069) | (0.073) |
| Kernel - PSM | 0.098 | -0.081 | -0.081 | 0.071 | 0.006 | -0.021 |
| | (0.083) | (0.103) | (0.112) | (0.055) | (0.071) | (0.078) |
| Nearest Neighbour | 0.319*** | -0.022 | -0.036 | 0.167*** | -0.022 | -0.055 |
| | (0.092) | (0.144) | (0.129) | (0.061) | (0.095) | (0.092) |

Note: Standard errors are reported in parentheses. In particular, robust standard errors (clustered by tax center) are reported for CEM estimator, bootstrapped standard errors based on 200 replications are reported for Kernel - MHD and Kernel - PSM estimators and heteroskedasticity-consistent analytical standard errors proposed by Abadie and Imbens (2006, 2008) are reported for Nearest Neighbour estimator; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E.4: Main Results – $ATT$ by audit type (trimming top 95% of treated and matched controls)

| Dependent Variable | CTI reported | | | CIT payable reported | | |
|---|---|---|---|---|---|---|
| Years after the audit | I | II | III | I | II | III |
| Type of Audit | (1) | (2) | (3) | (4) | (5) | (6) |
| Comprehensive | 0.364*** | 0.243 | 0.074 | 0.275** | 0.203 | 0.083 |
| | (0.137) | (0.207) | (0.219) | (0.118) | (0.182) | (0.171) |
| Narrow-scope | 0.073*** | -0.168*** | -0.105*** | 0.013 | -0.081* | -0.065* |
| | (0.023) | (0.056) | (0.027) | (0.016) | (0.042) | (0.036) |

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

## E.3: Placebo test

Our sample selected before the CEM stratification step (see Table 5) comprises of 100 businesses that will be audited for the first time in 2016 (out of the 337 taxpayers audited in the 2016 wave). To validate our matching methods we perform a falsification test by treating these businesses as they were audited in the 2015 wave and estimate the impact of this placebo enforcement on their 2016 tax reporting behaviour. We are able to find matches for 89 of them. Nevertheless, since these audits were initiated after the 2016 (but before the 2017) tax return was filed, their expected deterrence impact in 2016 should be equal to zero since the taxpayer would not have been aware of the audit when that return was filed. The estimated impact of this placebo treatment on the aggregate (see Table E.5, Columns 1 and 3) and for each audit type (Table E.6, Columns 1 and 3) is in fact small and statistically insignificant, consistent with expectations. By properly estimating the actual impact of audits one year after the treatment for this subset of businesses audited in 2016 and belonging to our matched sample, we find evidence of an aggregate pro-deterrence effect on their reporting behaviour. The estimation of the actual impact of 2016 wave of audit by audit type tend to confirm both qualitatively and quantitatively previous results on comprehensive audits while leading to inconclusive results for narrow-scope audits.[74]

Table E.5: Main Results – Aggregate $ATT$ (placebo 2016)

| Dependent Variable | CTI reported | | CIT payable reported | |
|---|---|---|---|---|
| Years after the audit | **Placebo** | **Actual I** | **Placebo** | **Actual I** |
| Matching estimator | (1) | (2) | (3) | (4) |
| CEM | -0.029 | 0.345*** | -0.044 | 0.307*** |
| | (0.069) | (0.133) | (0.041) | (0.092) |
| Kernel - MHD | 0.051 | 0.378* | 0.084 | 0.309* |
| | (0.219) | (0.204) | (0.166) | (0.163) |
| Kernel - PSM | 0.031 | 0.297 | 0.070 | 0.252* |
| | (0.207) | (0.194) | (0.158) | (0.150) |
| Nearest Neighbour | 0.127 | 0.381* | 0.067 | 0.280* |
| | (0.279) | (0.204) | (0.212) | (0.156) |

Note: Standard errors are reported in parentheses. In particular, robust standard errors (clustered by tax center) are reported for CEM estimator, bootstrapped standard errors based on 200 replications are reported for Kernel - MHD and Kernel - PSM estimators and heteroskedasticity-consistent analytical standard errors proposed by Abadie and Imbens (2006, 2008) are reported for Nearest Neighbour estimator; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[74] Notice that here we are estimating the impact of audit wave 2016 only on a subset of 89 out of 337 audited taxpayers and only for the purpose of the falsification test.

Table E.6: Main Results – $ATT$ by audit type (placebo 2016)

| Dependent Variable | CTI reported | | CIT payable reported | |
|---|---|---|---|---|
| Years after the audit | **Placebo** | **Actual I** | **Placebo** | **Actual I** |
| Type of Audit | (1) | (2) | (3) | (4) |
| Comprehensive | 0.109 | 0.285*** | 0.070 | 0.297*** |
| | (0.595) | (0.089) | (0.464) | (0.071) |
| Narrow-scope | -0.072 | -0.007 | -0.004 | 0.050 |
| | (0.197) | (0.212) | (0.178) | (0.144) |

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

## References

Abadie, A., Imbens, G.W., 2006. Large Sample Properties of Matching Estimators for Average Treatment Effects. Econometrica 74, 235–267.

Abadie, A., Imbens, G.W., 2008. On the Failure of the Bootstrap for Matching Estimators. Econometrica 76, 1537–1557.

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.

Armstrong, C., Glaeser, S., Huang, S., Taylor, D., 2019. The Economics of Managerial Taxes and Corporate Risk-Taking. The Accounting Review, 94, 1–24.

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.

Besley, T., Ilzetzki, E., Persson, T., 2013. Weak States and Steady States: The Dynamics of Fiscal Capacity. American Economic Journal: Macroeconomics, 5, 205–235.

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.

Boning, W.C., Guyton, J., Ronald H. Hodge, I., Slemrod, J., Troiano, U., 2018. Heard it Through the Grapevine: Direct and Network Effects of a Tax Enforcement Field Experiment. Mimeo.

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.

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.

Gamannossi degl'Innocenti, D., Rablen, M.D., 2020. Tax Evasion on a Social Network. Journal of Economic Behavior & Organization 169, 79–91.

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.

Gemmell, N., Ratto, M., 2012. Behavioral Responses to Taxpayer Audits: Evidence From Random Taxpayer Inquiries. National Tax Journal, 65, 33–57.

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

Hanlon, M., Mills, L., Slemrod, J., 2007. An Empirical Examination of Corporate Tax Noncompliance, in: Auerbach, A.J., Hines, Jr., J.R., Slemrod, J. (Eds.), Taxing Corporate Income in the 21st Century. Cambridge University Press, pp. 171—-210.

Hanlon, M., Slemrod, J., 2009. What does tax aggressiveness signal? Evidence from stock price reactions to news about tax shelter involvement. Journal of Public Economics, 93, 126–141.

Hasan, I., Hoi, C.S., Wu, Q., Zhang, H., 2017. Does Social Capital Matter in Corporate Decisions? Evidence from Corporate Tax Avoidance. Journal of Accounting Research, 55, 629–668.

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.

Heckman, J.J., Smith, J.A., 1995. Assessing the Case for Social Experiments. Journal of Economic Perspectives, 9, 85–110.

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.

Imai, K., King, G., Stuart, E., 2008. Misunderstandings Among Experimentalists and Observationalists about Causal Inference. Journal of the Royal Statistical Society, Series A, 171, 481–502.

Imbens, G., 2004. Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. Review of Economics and Statistics 86, 4—-29.

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

IMF, 2015. Current Challenges in Revenue Mobilization - Improving Tax Compliance. IMF Policy Paper, 5.

Jacob, M., Rohlfing-Bastian, A., Sandner, K., 2021. Why Do Not All Firms Engage in Tax Avoidance? Review of Managerial Science, 15, 459–495.

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

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., 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., Raaum, O., Torsvik, G., Øvrum, A., 2019. Compliance Effects of Risk-Based Tax Audits. CESifo Working Paper, 7616.

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., 2019. Active Ghosts: Nil-Filing in Rwanda. ICTD Working Paper, 106.

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, 2006. Strengthening Tax Audits Capabilities: General Principles and Approaches. Centre for Tax Policy and Administration Report.

OECD, AUC, ATAF, 2020. Revenue Statistics in Africa 2020. OECD Report.

Pomeranz, D., 2015. No Taxation Without Information: Deterrence and Self-Enforcement in the Value Added Tax. American Economic Review, 105, 2539–2569.

PwC, 2019. Worldwide Tax Summaries - Rwanda. Technical 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.

Rosenbaum, P.R., Rubin, D.B., 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 70, 41–55.

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

RRA, 2020. Tax statistics in Rwanda, Fiscal Year 2018/2019. Third Edition, June 2020, Kigali.

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

Slemrod, J., Gillitzer, C., 2014. Editor's Choice Insights from a Tax-systems Perspective. CESifo Economic Studies, 60, 1–31.

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.

UN-DP, 2020. COVID-19 and Human Development: Assessing the Crisis, Envisioning the Recovery. United Nations Development Programme, New York.

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.