THE IMPACT OF TAX AUDITS ON SMALL ENTERPRISES: COMPLIANCE GAINS OR BACKFIRING?

Christos Genakos[a], Christos Kotsogiannis[b], Luca Salvadori[c]

This version: February 5, 2025

Abstract: This paper investigates whether tax audits encourage compliance among small enterprises, which are numerous and challenging to monitor. Using administrative data from Greece, it is shown that audits can backfire, especially for noncompliant firms. Compliance responses vary by firm and audit characteristics, including inspection type, information source, and audited period. Full and partial audits show similar effects, while audits mandated by external to the revenue authority sources have little impact on compliance. Audits improve compliance when they are ongoing, at the time of filling. The return on investment for audits is positive and driven by verified underreported income.

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

JEL classification: H25, H26, H32.

[a]Cambridge Judge Business School, University of Cambridge, CEP and CEPR.

[b]Tax Administration Research Centre (TARC), Department of Economics, University of Exeter Business School, United Kingdom; and CESIfo, Germany.

[c]Autonomous University of Barcelona (UAB), Barcelona School of Economics (BSE), Spain; Tax Administration Research Centre - TARC (University of Exeter), United Kingdom; and Barcelona Institute of Economics (IEB), Spain.

Acknowledgments: We thank many officials of the Independent Authority for Public Revenue of the Hellenic Republic (IAPR) who have provided extensive feedback through insightful discussions. We also thank Andrea Viñas for excellent research assistance. Kotsogiannis acknowledges research funding from the Economic and Social Research Council (ESRC) (Grants ES/S00713X/1 and ES/X003973/1). Salvadori also acknowledges support from the Fundación Ramón Areces, the Serra Húnter Programme, the Generalitat de Catalunya [2021SGR00571], and the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S) and the grant PID2022-137707NB-I00 funded by MICIU/AEI/10.13039/501100011033 and FEDER, UE. Views expressed in this paper do not necessarily reflect those of the IAPR and its Management. The use of IAPR data in this work does not imply the endorsement of IAPR or its Management in relation to the interpretation or analysis of the information. The usual caveat applies.

1 Introduction

Tax evasion deprives the state of revenue, creates public deficits, distorts the tax burden on citizens, harms the welfare state, distorts the conditions of free market competition and leads to a lack of trust in institutions and State mechanisms on the part of taxpayers. Among the policy reforms for improving the efficiency of the tax systems (both in terms of the resource cost of raising taxes and the distribution of the tax burden) has been the strengthening of tax compliance capacity, an issue which has come to the fore for many countries around the world, following the financial crisis of 2008 and facilitated by the expanding access of academics to administrative tax-return data.[1] The safeguarding of tax compliance requires not only sufficient resources to the administration but also a compliance strategy which verifies that the true income has been reported by the taxpayer. Operational tax audits—defined broadly to include the selection and assessment of tax returns by the tax administration[2]—is, arguably, the key instrument in supporting tax compliance, and building capacity in this has been seen by many countries as the way forward to deliver more efficient and distributional fairer tax systems.

There is a growing literature discussing and evaluating the role of operational tax audits in tax compliance focusing, in particular, on whether they are conducive to compliance. Most studies (generally for advanced economies) find that there is a positive compliance impact on the aggregate, but one that varies both in magnitude across countries and across types of firms (Slemrod and Gillitzer, 2014). But what is also clear is that sweeping generalisations about the effectiveness of tax audits in promoting compliance is not advisable, due to the complexity inherent in these policy instruments and the challenges involved in designing and implementing an effective audit strategy (Slemrod, 2019).

There is also some evidence that tax noncompliance is related to observable characteristics of the firm showing that, for example, firms operating in cash-prevalent industries are more likely to be noncompliant, while third-party reporting reduces noncompliance (see, for example, Kleven et al., 2011; Advani, 2022; Christiansen, 2024). Related to this, there is also some evidence that certain economic sectors are more prone to being noncompliant (Artavanis, Morse and Tsoutsoura, 2016); as are small firms, more broadly (Slemrod and

[1] Taxation accounts for more than 80 percent of total government revenue in about half of the world's countries, and more than 50 percent in almost every country. This revenue is essential for the provision of public goods the support of the redistributive role of the state, and the mitigation of economic inequalities.

[2] Other important elements of the compliance capacity include procedures, penalties—not to mention of course the complexity of the tax system which might exacerbate noncompliance.

Gillitzer, 2014). But this evidence is scant.

Tax compliance is also likely to be related to elements that are internal to the Revenue Authority's policies such as, for example, the information conveyed to the taxpayer during the audit, the cause that has triggered the audit, and the frequency with which the firm is audited (a measure of audit intensity). To the best of our knowledge, and despite the flourishing literature analyzing tax compliance matters, there is a substantial lack of evidence regarding these issues. While tax data reflect not only the reporting behavior of the firms but also the enforcement behavior of the Revenue Authority (Slemrod and Gillitzer, 2014), investigating these additional margins provide arguably a better understanding of what matters in firms' tax compliance behavior.

The objective of this paper is to contribute to the understanding of the broader issue of tax compliance, focusing, in particular, on whether operational audits foster compliance among small enterprises (SEs), a subset of the broader category of small and medium firms (SMEs). These firms, for most advanced countries, account for well over 90 percent of all businesses and contribute to more than 60 percent of GDP (ILO, 2019).[3] In general, these firms are less likely to be audited despite of the fact that they tend to present a less compliant profile (Bachas, Fattal Jaef and Jensen, 2019).[4] In the U.S, for example, small firms that are audited infrequently show more noncompliance, as are those whose income is not subject to neither information reporting nor withholding (Slemrod and Gillitzer, 2014). In the UK too the estimated tax gap (the difference between the theoretical tax liability and actual collection) for SMEs account for just over 70 percent of the total tax gap (see HMRC, 2024).[5] For the purpose of this paper, SEs are defined as firms whose annual turnover does not exceed €10 million.[6]

This paper focuses on the role of two types of heterogeneities in determining the impact of audit deterrence. First, it examines the heterogeneity in terms of firm characteristics, where it investigates whether size, sector, compliance status, and the time of filling the tax return (relative to the outcome of the audit being known), matter for compliance. Second,

[3] In the EU small firms (that includes micro-sized too) accounts for over 98 percent of the total population.

[4] This presumably is driven by some form of optimal allocation of administrative resources. When revenue collections are weighed against administrative and compliance costs, the optimal policy would call for a particular focus on the largest payers, Keen and Slemrod (2017).

[5] In particular the share of the tax gap attributed to SEs alone represents the largest component of the tax gap and has increased steadily over the last 5 years, from 44 percent of the overall tax gap in 2018-2019 to 60 percent in 2022-2023.

[6] This classification follows the definition in Article 2 of the Commission Recommendation 2003/361/EC.

it analyzes the heterogeneity of the tax audit process related to information that initiated the audit; the number of inspections received in a year, the information the taxpayer receives during the audit, whether the audit is on-the-spot or desk-based and the audited period. Within this broad set of heterogeneities, the analysis estimates the compliance impact of the audit one year past the tax investigation, focusing on tax liabilities as the variable of interest.

To address these issues, use is made of a unique administrative data for the period 2012-2019 which consists of tax filings and audits performed by the Independent Authority for Public Revenue (IAPR) in Greece. More specifically, the analysis uses the universe of anonymized firm tax filings for the period 2012-2019 and the 2018 audit wave, as well as information on the risk rules and criteria employed by the revenue authority to prioritize tax filers for audit selection. Although the total number of audits is calculated centrally, taking into account the auditing capacity of the Revenue Authority, audits are performed locally, at the local tax office (LTO) to which the firms belong. Critically for us, while LTOs are assigned specific audit targets, they are also subject to many operational constraints, including the capacity of tax auditors and the complexity of the audit process, that vary in an unpredictable manner. In practice, this means that within a given LTO comparable enterprises might face differing outcomes, with some being audited while others are not. Likewise, two firms in different LTOs but with very similar characteristics, including their risk score, might be audited in one tax office but not in another. It is precisely these margins that are explored for the construction of counterfactual comparisons for each audited firm and in the empirical application of a matched Difference-in-Difference framework (matched-DID).

The results suggest the existence of a significant and robust counter deterrence ("back-firing") aggregate impact of tax audits, as SEs respond negatively by declaring less taxable income the year after the inspection. This is in contrast to the existing literature that typically finds a positive, but short-lived effect of tax audits. What drives this negative result in the present environment is the reaction of those firms who have been found non-compliant during the audit. Moreover, the larger the noncompliance identified during the audit or the tax burden (liabilities divided by taxable income), the less taxable income SEs declare the year after the audit. In contrast, there are no significant differences across SEs of different size, or across SEs in different sectors of the economy. In other words, the immediate backlash from noncompliant SEs that is documented here (and to the best

of our knowledge for the first time) seems to be different from the general tax evasion behavior known in the literature, as it is not exactly triggered by the same factors.

In terms of audit process heterogeneity, the evidence suggest a negative effect on compliance solely attributable to internal (versus externally mandated) audits, while it makes no statistical difference whether the audit was full (broad and exhaustive) or partial (narrow and focused). The year of tax liability that an audit is based on is important, as audits focusing on past years tend to have a more negative impact in terms of noncompliance. In addition, the counter deterrence effect of audits is stronger for firms audited two or more times in the same wave with respect to the others. Interestingly, the results also show that firms already undergoing an audit, but not yet informed about the outcome, declare higher tax liabilities the following year. Better understanding of such margins will not of course resolve all the design issues in tax compliance. But it can surely help in facilitating better guidance for the design of compliance strategies.

The rest 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. Section 3 lay out some conceptual considerations underlying the impact of tax audits on SEs compliance behavior, while Section 4 presents the institutional setting. Section 5 describes the data and the methodological approach followed. Section 6 presents the results regarding the aggregate impacts and along the heterogeneity elements analyzed, and also provides a back-of-the-envelope calculation on the returns to tax auditing. Section 7 summarizes and provides some concluding remarks.

2 Literature review

The literature in this area is growing in research contributions and policy documents, with the latter providing generic information and advice to tax authorities on tax audits. This growing interest reflects not only the importance of the issue, and the academic and policy interest in this area, but also the idiosyncrasies that exists across countries (institutionally but also in terms of tax systems) which make policy generalizations an impossible task. There is, however, some commonality in the findings, with most of the recent contributions[7] finding an aggregate significant average positive impact of tax audits

[7]Mostly focusing on personal income taxes in high income countries: See, for example, Kleven et al. (2011), Løyland and Øvrum (2017), DeBacker et al. (2018*a*), DeBacker et al. (2018*b*), Beer et al. (2020), Advani, Elming and Shaw, 2021 and Boning et al. (2024).

on the future reporting behavior of audited taxpayers. Perhaps less immediately evident is the question of whether tax audits might backfire thereby increasing noncompliance. Arguably, one can think of instances where this might be the case when, in particular, audit intensity signals distrust in taxpayers thereby generating the perception that the Revenue Authority and its enforcement actions are excessive and unfair (as in, for example, Kirchler, Hoelzl and Wahl, 2008; Wahl, Kastlunger and Kirchler, 2010). More recently, the literature has focused on evaluating the impact of tax audits in corporate taxes (Kotsogiannis et al., 2024) and VAT (Best, Shah and Waseem, 2021; Kotsogiannis et al., 2025; Henning et al., 2024), with the evidence too suggesting that in some instances tax audits do have the undesirable property of non-increasing compliance in the future. The challenge therefore is to identify which elements of the audit deliver (or not) compliance, an issue that is important for tax compliance design.

This paper focuses on corporate income taxes and SEs, paying particular attention to areas where limited evidence exists. For example, by focusing on SEs, it sheds light on whether audits trigger different response depending on the tax base and, more precisely, the type of taxpayer (see Slemrod and Gillitzer (2014) for a discussion on this issue). It also examines several relatively unexplored heterogeneous effects such as the type of the inspection (on-the-spot or desk-based), whether the taxpayer has been flagged internally (by the Revenue Authority) or externally (by other government departments that can legally do so), the intensity of the inspection in terms of audit frequency, the age of the audit case, and the type of information received by the taxpayer relative to the submission of the tax assessment. There are other dimensions of interest too, such as the sector of their main activity, their assessed (non-)compliance and their size. Despite the sizeable (and growing) literature analyzing tax compliance issues, there is a lack of evidence regarding the differential impact of audits on these margins.[8] The main contribution of this paper is to shed light on these aspects. Indeed, as noted earlier, the existing literature indicates a positive—but predominantly short-lived effect of audits—with a significant impact observed in the initial years after the audit, followed by a return to pre-audit compliance levels (see, for example, Kleven et al., 2011; Løyland and Øvrum, 2017; DeBacker et al., 2018*a*,*b*; Beer et al., 2020; Advani, Elming and Shaw, 2021).[9] Some studies have explored

[8]Some evidence on few of these margins but for developing countries can be found in Best, Shah and Waseem (2021), Kotsogiannis et al. (2024), Kotsogiannis et al. (2025), and Henning et al. (2024).

[9]This effect is primarily driven by self-reported income (see, for example, Kleven et al., 2011 and DeBacker et al., 2018*a*). The impact of audits might be significant and sustained. Boning et al. (2024)

various dimensions, such as the sector, engagement in exports/imports, legal type, and age of the business, to shed light on potential heterogeneities in the deterrence effect of audits. However, research on the specific deterrence effect on these margins remains inconclusive (Gemmell and Ratto, 2012; Løyland and Øvrum, 2017; DeBacker et al., 2018*a*; Beer et al., 2020) and several margins unexplored.

This paper also contributes to several additional strands of the literature on tax compliance. First, it contributes to the ongoing research on policy responses to tax evasion (reviewed in Andreoni, Erard and Feinstein, 1998; Slemrod and Gillitzer, 2014; Slemrod, 2019). A core part of the emphasis is on the role of information, reinforcing results from studies investigating the importance of third-party information reporting on tax compliance (see, for example, Mittal and Mahajan, 2017; Slemrod et al., 2017; Naritomi, 2019; Pomeranz and Vila-Belda, 2019; Brockmeyer and Sáenz Somarriba, 2025). Importantly, the analysis here also relates to the impact of audit probability on taxpayer behavior (see Slemrod, Blumenthal and Christian, 2001; Alm and Malézieux, 2021), with the deterrent effect deriving from various aspects of the audit process. Leveraging novel information provided by auditors on taxpayers' perceived willingness to comply, Christiansen (2024) finds that unintentional noncompliers, driven by inattention or misunderstandings of the tax rules, exhibit higher compliance in subsequent years. In contrast, intentional noncompliers who deliberately evade taxes and are typically targeted for operational audits do not respond to audits and have a low recovery rate for evaded taxes.

There are important policy prescriptions emerging from the analysis. Tax audits on SEs are crucial for ensuring compliance with tax laws and maintaining the integrity of the tax system, ensuring that all firms contribute their fair share to public finances. However, if not conducted carefully, audits—as the analysis will show—can lead to an undesirable outcome regarding compliance. The next section provides a conceptual framework whose sole purpose is to highlight the mechanism underlying the empirical results.

find that tax audits generate a sustained increase in future tax payments resulting in additional taxes that are 3.2 times higher than the tax initially uncovered. Business income allows more leeway for taxpayers to alter reported income from year to year, contributing to less persistent effects compared to labor income. The sophistication level of taxpayers also matters, with more experienced filers being less affected by tax enforcement (DeBacker et al., 2018*a*).

3 Conceptual considerations

The model[10] is intentionally simple and its sole purpose is to identify an ambiguity that is central to the issues discussed in this paper. Consider a firm that produces a homogeneous output q, at cost $\psi(q)$ (with $\psi'(q) > 0$), which is sold at price p. The firm pays tax on the sales of the good pq,[11] but saves on taxes if they underreport an amount of some e. Underreporting of sales is costly, with the cost being given by $c(e, \alpha)$ with $c_e(e, \alpha) > 0$;[12] where α is a (continuous) variable that captures the intensity of auditing of the Revenue Authority. The cost of underreporting determines a firm's underreporting level, and it might be conceivably the case that, given intensity of auditing α, the cost of underreporting is high (low) and so firms' underreporting is low (high). Importantly, the cost of noncompliance is also affected by the intensity of auditing.[13]

The precise shape of $c(e, \alpha)$ is therefore an empirical question. Intuition suggests that the cost of noncompliance is high if enforcement processes (from prioritization to penalties for underreporting to the identification of underreporting) are effective. Imagine, for example, that enforcement of the Revenue Authority focuses rather mechanically on a proportion of the tax base, say VAT refunds, to give an example and by no means the only. What is the impact of such mechanical audit on compliance following the audit, if such a mechanical audit does not discover the extent of underreporting across all tax bases the taxpayer might be underreporting?[14] It is so conceivable that the impact of the audit might be to reduce the (realized) cost of evasion, thereby increasing evasion. Arguably, this is not an appealing feature of the audit, yet it has been observed in recent empirical studies that use tax administrative data.[15] It is this point that this simple model wishes

[10] A variant of Keen and Slemrod (2017), and departs, for example, from considering market structure and details of the enforcement policy of the Revenue Authority. All these are important aspects, some of which are discussed in Bayer and Cowell (2009).

[11] Introducing tax deductible expenses does not add any additional insights.

[12] Evading sales is costly for firms, both in time needed to find the best avenues to evade but also (on expectation) the fine the firm needs to pay if caught evading.

[13] This is an important function, and can be derived from first-principles introducing beliefs: for an approach along these lines see Kotsogiannis et al. (2024).

[14] As Slemrod and Gillitzer (2014) notably put it, during an examination '[s]ome noncompliance maybe missed, and there will also be mistakes in characterizing as noncompliance what is legitimate tax planning,' p.44. The outcome of the tax audit might also depend on the duration of the audit, and therefore 'the tax liability per the originally filed return... maybe partly a tactical "opening bid" that is neither party's best estimate of the 'true" tax liability, Slemrod and Gillitzer (2014), p.45. Similarly, resource constraints and audit procedures can limit the comprehensiveness of the examination of noncompliance.

[15] See, for example, Erard, Kirchler and Olsen (2019), Best, Shah and Waseem (2021), and Kotsogiannis

to emphasize.

The firm's objective function is to maximize

$$\pi\left(q,e;t,\alpha\right)=pq-t\left(pq-e\right)-c\left(e,\alpha\right)-\psi\left(q\right), \tag{1}$$

with necessary conditions given by

$$\pi_q=\left(1-t\right)p-\psi\left(q\right)=0, \tag{2}$$

and

$$\pi_e=t-c_e\left(e,\alpha\right)=0. \tag{3}$$

Equation (3) implicitly defines the optimal level of underreporting as a function of the parameters, $e\left(\alpha,t\right)$, with, in particular,[16]

$$e_\alpha\left(\alpha,t\right)=-\frac{c_{e\alpha}\left(\alpha,t\right)}{c_{ee}\left(\alpha,t\right)}. \tag{4}$$

It is the sign of equation (4) that is the focus here. With the sufficiency condition requiring $c_{ee}<0$, it is the case that the sign of $e_\alpha\left(\alpha,t\right)$ takes the sign of $c_{e\alpha}\left(\alpha,t\right)$.

As already noted, there is no a priori reason why more intensity in auditing improves the accuracy of the audit process, so it is conceivable that c_{ea} is positive (or not very negative).[17] In a more general level, the (sign of) function c_{ea} reflects the subtlety of the audit process. A tax audit, for example, might create uncertainty: due to the complexity of the tax law what is actually tax liability is often unclear. Any tax compliance audit is not perfect, and some noncompliance might be missed (unintentionally, or intentionally), as there would be mistakes in assessment (Slemrod and Gillitzer, 2014; Best, Shah and Waseem, 2021; Kotsogiannis et al., 2024).

A clarification is in order here. First, as emphasized already, the impact of tax audits is not homogeneous across all firms. Heterogeneities which can affect the audit outcomes arise from various factors such as the size and the sector in which a firm operates. Small

et al. (2024). As noted in Section 2, there are many reasons why tax audits might backfire, ranging from taxpayers inferring that tax evasion is more widespread than they previously thought to their perception that tax audits are excessive and unfair.

[16] A consequence of the specification is that the decision of how much output to produce is independent of the level of evasion.

[17] The underlying assumption behind the analysis of Keen and Slemrod (2017) is that more capacity improves the accuracy of the audit process—this might not be the case if the auditing process focuses on examinations that are mechanical, for example.

firms may face different challenges and incentives compared to large ones. It has been argued that small firms, when at a competitive disadvantage, are more likely to engage in tax noncompliance (Cai and Liu, 2009; Wang, 2012). Firms with prior audit experience may also respond differently compared to those being audited for the first time, or firms who are being audited during the time of filling. Secondly (and related to the point just made), the intensity of the audit is not homogeneous across all firms, and the information gathered by the revenue authority may be internal (leveraging their data) or external, and coming from third party sources, and can be mandated. Though the model abstracts from these issues, they are nevertheless aspects that the analysis will explore.

4 Institutional background

In 2010 the inability of the Greek government to borrow funds from international markets led to a €110 billion bailout loan from the European Commission, the European Central Bank, and the International Monetary Fund. As part of the loan agreement, the Greek government agreed and implemented a series of structural reforms with the aim to modernize and digitize the public sector. One of the most significant and emblematic reforms was that in 2016 the Revenue Authority became functionally independent with administrative and financial autonomy, a reform that was deemed necessary for the organization to become more flexible and efficient. The new Revenue Authority, shielded from political influences, and with the support of the international organizations participating in the bailout program, quickly started adopting best practices from other tax administrations.

One of the main focus in this process has been the improvement of risk management of tax audits—which includes the identification, evaluation, analysis, measurement, and interpretation of risks—with the aim of reducing the possibility of tax revenue losses from the failure to prioritize appropriately and carry out effective tax audits.[18] Considerable effort has been put into developing further the automated process of evaluation of risks making sure that there is no possibility of human interference once the key parameters (criteria and weights) have been determined. And this research—in collaboration with the IAPR and having access both to the risk score given to small firms and the list of criteria used—is part of the broader strategy of the Revenue Authority to understand the

[18]Part of the reforms were targeted on reducing the tax gap and improving tax compliance. The tax burden (tax revenues as percentage of GDP) in 2018 and 2019 in Greece was 42.7 percent and 43.6 percent, respectively, above the European Union average of 41.2 percent in 2018, and 41.1 percent in 2019.

effectiveness of tax audits. The time chosen is the year 2018, which is the year that the bailout program was officially concluded.

The audit prioritization in 2018 used the criteria developed in that year, but was applied to the economic years determined in the audit. The list of criteria used in the risk assessment is not made public, as are any other audit-related information. The checking of the fulfilment of tax obligations of the taxpayer is carried out by the tax administration, and it may take the form of a tax audit from the offices (desk-based), or it can take the form of on-the-spot (full) tax inspection. The tax administration may carry out a tax audit from its offices on the basis of financial statements; returns and other documents submitted by the taxpayer, as well as documents and information in its possession or on the basis of books and accounting records produced by the taxpayer. The decision of the tax administration to conduct a full on-the-spot tax inspection must be notified to the taxpayer with prior written notice. An on-the-spot tax inspection may be carried out without prior notice in cases where there are indications that tax evasion has been committed. The tax administration has the right to review a tax period or case for which a full audit has already been carried out only if new elements emerge from which the tax liability exceeds that determined following the initial check. There is no time limit on the conclusion of a tax audit, and it is possible that while an audit is ongoing, the firm receives limited information about the audit's outcome, even before filing the tax return following the audit.

The number of audits is calculated taking into account the auditing capacity of the Revenue Authority. In addition, a number of tax audits were mandated for investigation from sources external to the organization, such as the Financial Crime Investigation Directorate, the Public Prosecutor's Office and other relevant Judicial Authorities. Although all firms submit their tax assessment to the LTOs, for audit purposes, cases are allocated to audit services based on turnover. Specifically, firms are assigned to the LTOs' audit services if they report gross sales of no more than €10 million in the fiscal year. Audits are performed by the 57, geographically dispersed, LTOs to which the firm belongs. The LTOs are monitored by the Tax Regions, who are responsible for monitoring the correct and uniform application of the legislative provisions, circulars and instructions and their operational procedures, providing guidance to them through random or targeted checks on the correct application of tax and other legislation.[19] Critically for the analysis, although

[19] Article 43, Government Gazette 4,738 B'2020.

LTOs are given specific audit targets, various operational constraints (for example, personnel available, difficulty of the audit) in practice means that two comparable enterprises inside the same tax office can have one audited, while the other is not. Likewise, due to staffing and time constraints, two firms with very similar characteristics, including their risk score, maybe audited in one tax office but not in another. It is precisely these margins that we explore in our empirical application later by matching firms based on a number of factors to construct their counterfactual comparisons.

The analysis focuses on small enterprises (SEs) as defined by the IAPR (gross sales of no more than €10 million in the fiscal year) for two main reasons. First, SEs represent the backbone of the Greek economy, as they account for 94.6 percent of the number of businesses. Second, tax evasion is thought to be prevalent among them and really hard to control. This is not a phenomenon that applies only in Greece, but it has been highlighted in many advanced countries (see, for example, OECD, 2015, Chapter 4), both in the U.S (Slemrod and Gillitzer, 2014) and in Europe (for the UK see HMRC, 2024). The audit criteria are designed precisely with the aim to prioritize audits and save precious resources given the large number of SEs and, as part of the compliance strategy, will discipline the unaudited firms. In this respect, the analysis investigates not only the possible heterogeneity in terms of the firms' observable characteristics (such as industry, turnover size), but also likely heterogeneity and the efficacy of different dimensions of the audit process itself (the audited period, whether the audit is full or partial, whether the audit is ongoing while the legal entity files its tax return) for which there is limited (to no, for some dimensions) evidence in the literature.

5 Data and methodological approach

5.1 Data

To assess the impact of tax audits on SEs' reporting behavior, three datasets are combined containing data on income tax declarations, tax audits and audit process. The data on income tax declarations include all underlying raw data on income/expenditures inflows and outflows, as well as industry the firms belong to for all firms that were prioritized in 2018—that is, those that were scored by the revenue authority based on a number of objective criteria covering the tax years 2012 to 2018,[20]—and their total risk score,

[20] All data have been anonymized at the source by the IAPR and no individual SE can be identified.

as well as its individual components, and their audit priority number, provided by each LTO. The tax audits data contains the records for each of the SEs and whether they were audited in 2018. The audit process data has records of the various details regarding the audit process and, in particular, whether the audit was full or partial, the length of time under which a firm has been under a tax investigation, how many times a firm has been audited, and whether the tax audit has been mandated by an institution external to the Revenue Authority. Most importantly, the dataset contains information on whether the audit process was concluded or not (at the time of the data collection, one year and a half after) and its final outcome that is, whether the firm was found compliant or not and the total monetary amount it had to pay in the latter case. These firms have been matched with their tax declaration records one year after the audit process (2019).

Having access to such detailed administrative data allows us to explore the tax audit effectiveness both along firm characteristics but also across the different aspects of the audit process. The strategy to do so is to use the audited firms as the treatment sample and construct various control groups by matching the treated firms with similar characteristics (risk score, industry, geographical location and prioritization number) that were not audited and compare their tax declaration behavior one year after the audit process.

The data has undergone various cleaning procedures. Most importantly, SEs reporting missing or null taxable income, while possibly indicating positive liabilities, have been excluded. There are SEs which, despite having ceased their activities, have a tax record due to existing debts with the tax authorities. As there is no trading, and no income reported in 2019, these (11,227) firms have been eliminated from the data, allowing to focus the analysis on firms that are operationally active.[21] The final dataset consists of 5,132 firms (41,056 firm-years observations), distributed across the 57 LTOs. In terms of audits initiated in 2018, we observe 2,072 completed inspections[22] and 70 which are ongoing in the sense that the legal entity files its tax return while the outcome of the audit is not known. The analysis focuses on completed cases, but it will also compare the estimated

[21] As a robustness check, the analysis has been replicated with the inclusion of these firms and the results obtained are both qualitatively and quantitatively similar. In the interest of space, given the already lengthy Appendix, this robustness analysis is available upon request.

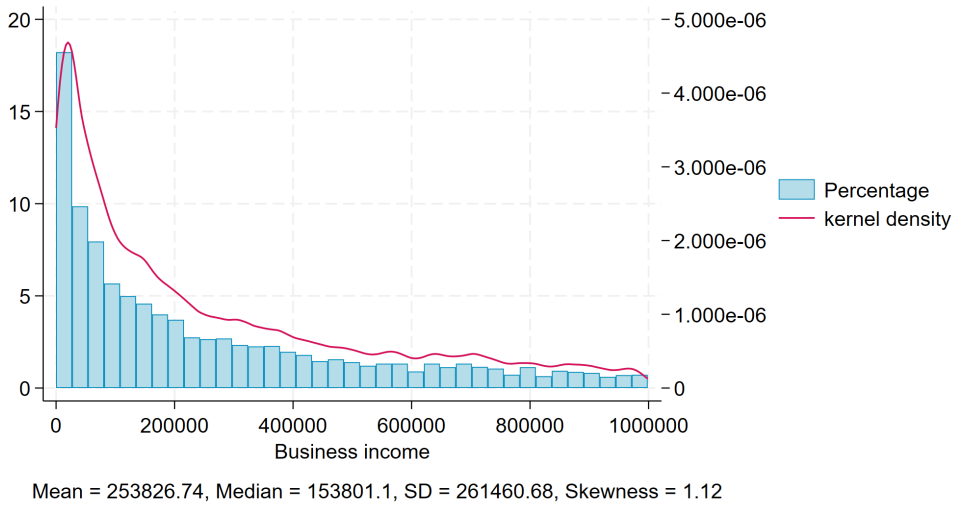
[22] If the taxpayer disagrees with the assessment they have the right to dispute by submitting an appeal to the Tax Dispute Unit. If the taxpayer subsequently disputes the unit's decision they have the right for an appeal to the administrative court for a final decision. The audits in the data are deemed complete and closed in the sense that the amounts verified are not subject to any dispute, and any dispute that has existed has been dealt with, with both sides agreeing to the final amount verified.

aggregate deterrence impact of these audits, and their detection power, with the impact of the ongoing ones. The following subsections provide some descriptive statistics on the final selected dataset of 5,132 firms and in particular some descriptive analysis on the completed audits.

Descriptive statistics: SEs characteristics

Figure 1 presents the distribution of SEs based on their gross income which is skewed to the right. On average SEs have a gross business income around €253,827, with a median of €153,801 and a standard deviation of €261,460. Around 60 percent of the SEs have a gross income below €200,000.

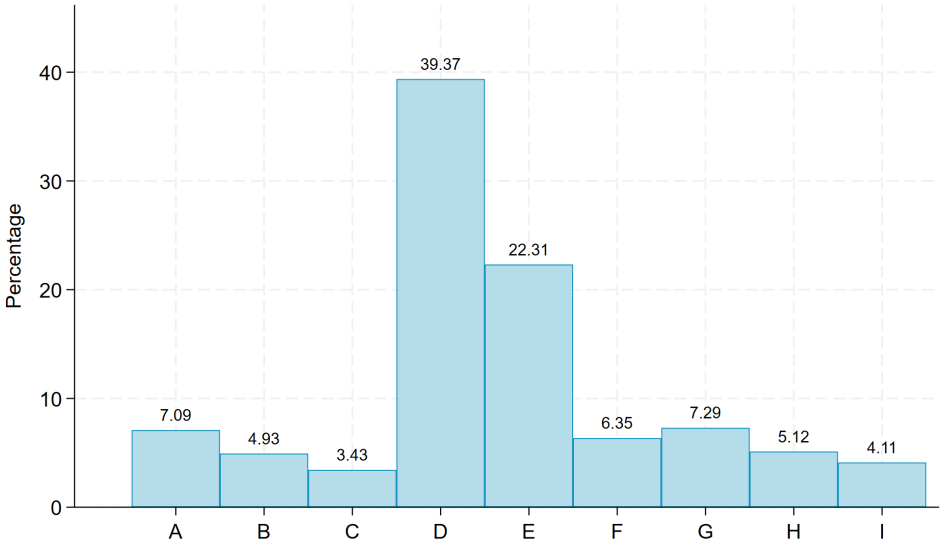
Figure 1: Distribution of SEs by Gross Income



Note: Authors' calculations based on the sample of SEs with completed audits.

Figure 2 reports the distribution of firms by Industry according to the ISIC classification. Most SEs in the sample tend to concentrate in Industry D that represents almost 40 percent of SEs operating in the core service and distribution industries (trade, logistics, and hospitality), followed by Industry E grouping around 22 percent of firms active in the information and communication sectors, while the remaining firms are uniformly spread across industries.

Figure 2: Distribution of SEs by Industry



Note: Authors' calculations. Industries are grouped according to the high-level ISIC aggregation: A=Agriculture, forestry and fishing; B=Manufacturing, mining and quarrying and other industry; C=Construction; D=Wholesale and retail trade, transportation and storage, accommodation and food service activities; E=Information and communication; F=Financial and insurance activities; G=Real estate activities; H Professional, scientific, technical, administration and support service activities; I=Public administration, defence, education, human health and social work activities.

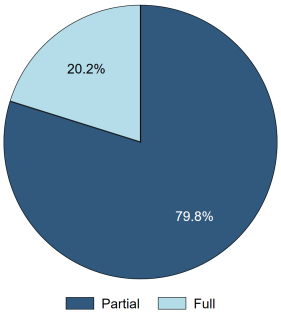
Descriptive statistics: Audit process characteristics

Distribution across audit categories Tax audits can be classified into different categories. First, they can be distinguished in terms of their type or scope, and categorized in full or comprehensive and partial or desk-based audits. As can be seen from Figure 3.a, the majority of tax audits conducted are partial audits (79.8 percent), with full audits constituting just over a fifth of the total. Secondly, tax audits can be categorized in terms of the origin of information used by the revenue authority for their selection, with most cases being selected by the revenue authority (internal) and accounting for 88.6 percent of all cases, whereas only 11.4 percent are mandated by other government bodies (external), as shown in Figure 3.b. Tax audits can be also clustered following the period of tax declarations an audit refers to relative to the audit year. Figure 3.c presents the distribution of tax audits based on the period of examined tax declarations relative to the 2018 audit year: for tax declarations up to 3 years old (2014-2016 and denoted as T_3), up to 5 years (2013-2017 and denoted as T_5), and those exceeding 5 years (before, and including, 2012: $T_{>5}$). Audits up to 3 years represent 34.4 percent of the total, up to 5 years 23.3 percent, and those exceeding 5 years 42.3 percent. Finally, audits can be categorized in terms of their intensity measured by the number of times a firm has been audited within the same audit wave. As can be seen from Figure 3.d, the large majority of audits refer to firms receiving a single inspection in 2018 (88.9 percent), while firms audited multiple times within the same audit wave represent 11.1 percent of the total.

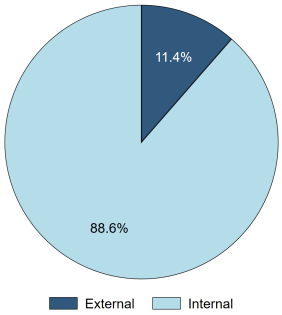
Compliance status Figure 4 depicts the distribution of audits by compliance status, both in the aggregate and across the categories of Figure 3. On the aggregate, the majority of cases correspond to taxpayers found compliant (63 percent) while the rest are determined noncompliant (37 percent). The success in uncovering noncompliance of an audit varies upon the type of audit, with full audits detecting 62.4 percent cases of noncompliance, while for partial audits this figure is halved (around 30 percent). Similarly, also across the audit category, inspections are not uniform in their noncompliance detection capacity with external audits uncovering noncompliance in the 65 percent of cases, while internal audits in 33.4 percent of cases. The distinction is less pronounced if the audits are categorized based on their seniority, with older cases detecting noncompliance more frequently than recent ones. The heterogeneity in detection power is also nuanced if audits are clustered by

Figure 3: Distribution of Audits Across Different Categories

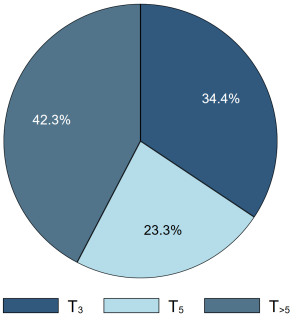
(a) Audit Type



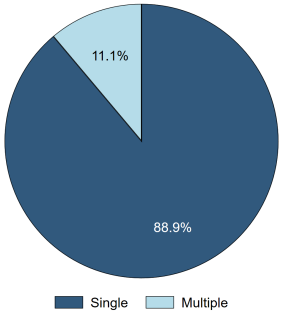
(b) Information Category



(c) Tax Declaration Period Relative to Audit Year



(d) Audit Intensity



Note: Authors' calculations.

their intensity, proxied by the number of years of tax declarations that have been audited during the 2018 audit wave.

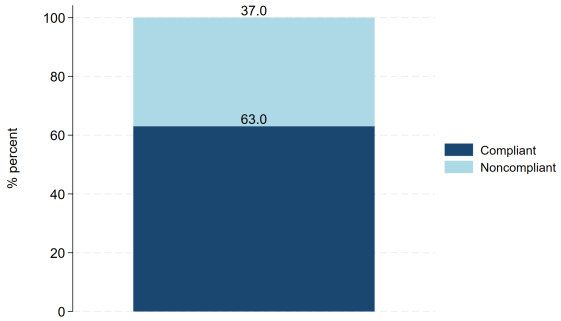
Audits by taxable income decile Figure 5 illustrates the distribution of audits by taxable income decile, both in the aggregate and by type of audits (Partial *versus* Full). The distribution appears uniform, with slightly more variability observed in the case of full audits.

5.2 Methodological approach

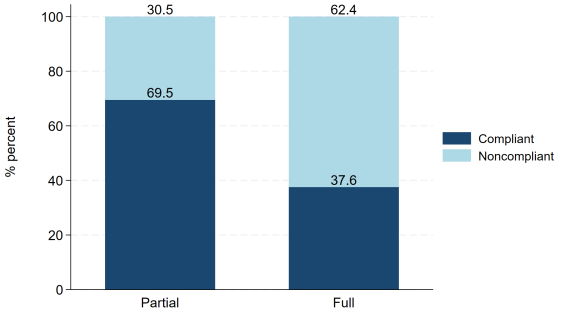
To estimate the impact of audits on future reporting behavior we combine a Difference-in-Difference (DID) approach with matching. This strategy aims at addressing the presence of a selection bias in the treatment assignment, like in the context of risk-based audit selection. Briefly, the basic idea behind matching is to pair each member of the treatment group with

Figure 4: Distribution of Audits by Compliance Status

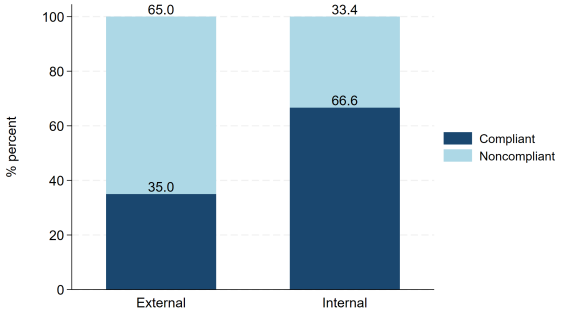
(a) Overall



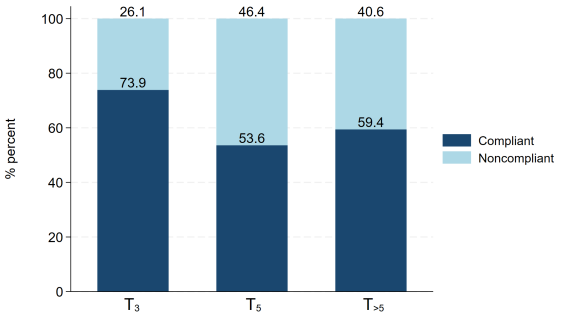
(b) Audit Type



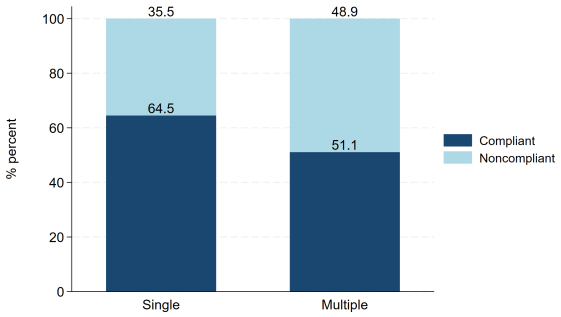
(c) Information Category



(d) Tax Declaration Period Relative to Audit Year



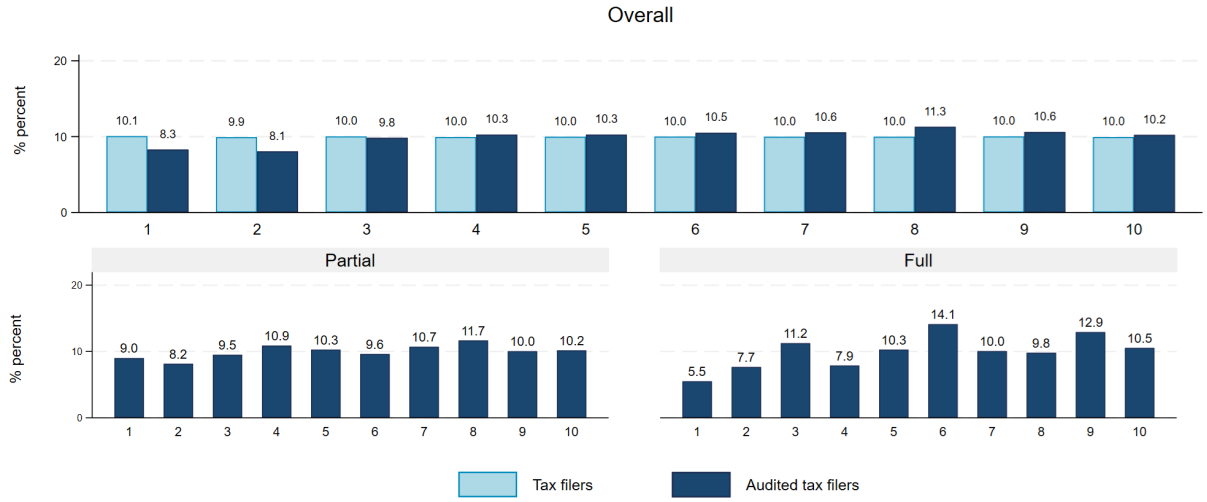
(e) Audit Intensity



Note: Authors' calculations.

a set of observationally equivalent control group members. By holding the confounding factors constant, the difference between the outcome variable of audited (treated) firms and matched controls (unaudited) is a direct estimate of the treatment effect that does not rely on any parametric assumptions.

Figure 5: Distribution of Taxpayers and Audits by Income Deciles



Note: Authors' calculations.

To be more precise, the analysis relies on Coarsened Exact Matching (CEM)[23] to stratify taxpaying firms based on relevant observable characteristics and utilizing a weighted regression model with weights resulting from that stratification providing balancing between treated and control units.[24] For the stratification, we use five progressively less parsimonious sets of pre-treatment covariates. The basic set of matching variables – Set I –includes the synthetic Risk Index calculated based on the risk criteria provided by the IAPR and the industry sector (according to the ISIC classification) that the SE belongs to. This set of variables synthetically accounts for taxpayer riskiness of underreporting their tax liabilities and the characteristics related to the sector of activity. Set II comprises Set I and adds the priority number, a ranking of cases specific to each LTOs. Set III consists of Set I plus LTO dummies, whereas Set IV consists of Set I and both the priority number and the LTO dummies. Finally, Set V matches on the ISIC industry and accounts for taxpayer riskiness in a more fine-grained manner by matching on all the audit selection criteria rather than the synthetic risk score. This approach varies from the one used in Set I in that firms are

[23] See Iacus, King and Porro, 2011, 2012, 2019.

[24] See Appendix A and B for more details on the methodology. See also Kotsogiannis et al., 2024, 2025 for a similar application of this strategy within this strand of the literature. For a discussion on alternative matching techniques, see, among many others, Abadie, 2005; Wooldridge, 2002, 2007; Stuart, 2010; Cattaneo, 2010; King et al., 2011; Imbens and Rubin, 2015; Guo and Fraser, 2015; Heckman, Ichimura and Todd, 1998.

matched based on their likelihood of being noncompliant according to the specific source of probable noncompliance rather than using a synthetic aggregate measure. In all cases, the outcome variable is the self-reported SEs' tax liabilities expressed in natural logarithms.[25] This is the most relevant margin because it measures the net impact on revenues collected by the Revenue Authority.

Using the weights resulting from this stratification, the estimated Average Treatment effect of audits on the reporting behavior of Treated (that is, the audited) taxpayers (ATT) can be obtained as

$$\widehat{ATT} = \frac{1}{N^1} \sum_{i:D_i=1} \left[\Delta Y_i^1 - \sum_{j:D_j=0} W(i,j)\, \Delta Y_j^0 \right], \tag{5}$$

where ΔY_i^1 is the post-audit change in SE i's reported outcome if the SE has been audited, ΔY_i^0 the analogous change if they were not audited, N^1 the number of treated SEs, and $\hat{C} = \sum_{j:D_j=0} W(i,j)\, \Delta Y_j^0$ the matching estimator providing an estimate of the counterfactual outcome by weighting the observations in the control group based on their similarities with the ones in the treatment group according to the CEM methodology and as synthesized through the weights $W(i,j)$. The estimated ATT is obtained through the weighted regression model

$$\Delta Y_i = \alpha + \beta \cdot audit_i + \gamma' \mathbf{R}_i + \epsilon_i, \tag{6}$$

where the impact of audit treatment ($audit_i$) is measured by the coefficient β, and therefore $\widehat{ATT} = \widehat{\beta}$, and ϵ_i is the error term. Estimation of (6) is performed with pre-treatment control variables,[26] included in the vector $\mathbf{R}_i$, to further reduce any residual imbalance and without (with the results of the latter being reported in Appendix C). As a robustness check, we use Propensity Score Matching (PSM) instead of CEM and estimate equation (6) with corresponding weights.[27] The results of this analysis are reported in Appendix D.

[25] SEs that have reported null tax liabilities are excluded.

[26] These include variables such as the risk score, the priority number and a set of dummies for the LTOs.

[27] In our context, PSM involves estimating the probability of being audited (the propensity score) through a logit model, where the dependent variable represents assignment to the audit treatment and the controls are the same variables that we employ for CEM, meaning that we estimate five alternative propensity scores based on the sets of variables used for CEM stratification. The counterfactual term is then estimated as a weighted average of outcomes, considering the similarity across taxpayers with respect to the estimated propensity score. We opt for the Kernel estimator within the PSM class, following Heckman, Ichimura and Todd (1997, 1998), as it provides more efficient weighted average counterfactuals based on all control group units.

The identifying assumption is that the outcome for audited SEs would have evolved in parallel to the outcome for control units in the absence of the audit. To test the presence of a common pre-treatment trend in the evolution of the outcome variable, we rely on Weighted Fixed Effect regressions based on weights obtained through our different matching approaches, with our event-study equation taking the form of

$$\Delta Y_{it} = \alpha + \sum_{k=-3}^{1} \beta_k \cdot audit_{it+k} + \theta_i + \tau_t + \varepsilon_{it}, \tag{7}$$

where ΔY_{it} is the change in i's reported outcome at time t from the $t-1$ period, $audit_{it+k}$ accounts for the audit treatment, β_k represents the period specific pre- and post-treatment effects, θ_i accounts for SEs fixed effects and τ_t accounts for time fixed effects controlling for specific economic cycles at year×ISIC industry×LTO level, while ε_{it} is the error term. By first differencing the data, we can take into account firm-specific trends in addition to firm-specific time-invariant factors. This approach largely addresses any remaining concerns related to the possibility of significant differences in SEs' trends before the audit treatment.

While the focus of the analysis is on closed cases—the large majority of audits in the sample–notably, we extend our investigation beyond this, allowing for a comparative analysis with tax audits that are ongoing at the time of the taxpayer reporting their tax liabilities. Furthermore, in an effort to elucidate the nuanced effects underlying the overall influence of closed audits on deterrence, our study delves into various dimensions, aiming to illuminate the origins of the estimated impact. As previously discussed, these dimensions can be conceptually categorized into two primary groups: heterogeneities in firm characteristics and heterogeneities in the audit process. Within the first cluster, we delve into the outcome of the audit, discerning between compliant and noncompliant audited taxpayers. Furthermore, we explore the quartiles of noncompliance to gain a more nuanced understanding. Additionally, we scrutinize the quartiles of the effective tax rate, serving as a measure of the tax liability's relative size, and the income quartiles, functioning as a proxy for business size. Finally, we investigate the impact across the ISIC industry classification. Regarding the second group of potential heterogeneities, the analysis assesses the impact based on operational dimensions linked to the audit process. As detailed in Section 4, the revenue authority employs diverse enforcement instruments, prompting a differentiation in the analysis based on the type of audit (Partial *versus* Full). As already

discussed in Section 5.1, within the audit landscape, distinctions also arise based on the origin of the information triggering the audit, leading to a categorization into external *versus* internal audits, an aspect subjected to our analytical scrutiny. The dataset allows us to further deepen our exploration by considering the differential impact of audits based on the year of the audited period relative to the 2018 audit wave, and the audit intensity (quantified by the number of times a firm undergoes auditing).

6 Results

This section is dedicated to presenting the results, commencing with an exploration of the overall ATT of closed audits. Following this, we delve into the nuances through heterogeneity analyses related to the firm characteristics and the audit process itself. These analyses aim to uncover variations in treatment effects across different dimensions (internal and external to the firm), providing a more granular understanding of the SEs' compliance responses. We also compare the ATT of closed cases with that of ongoing audits. The section then concludes with a back-of-the-envelope calculation of the Return on Investment (ROI) stemming from the audit interventions. Additional results are reported in Appendix C, presenting the baseline CEM matched-DID models without controls. To ensure the robustness of our primary results and methodology, sensitivity analysis is provided in Appendix D, offering a comprehensive validation of all the findings presented in the main text using PSM as an alternative matching method.

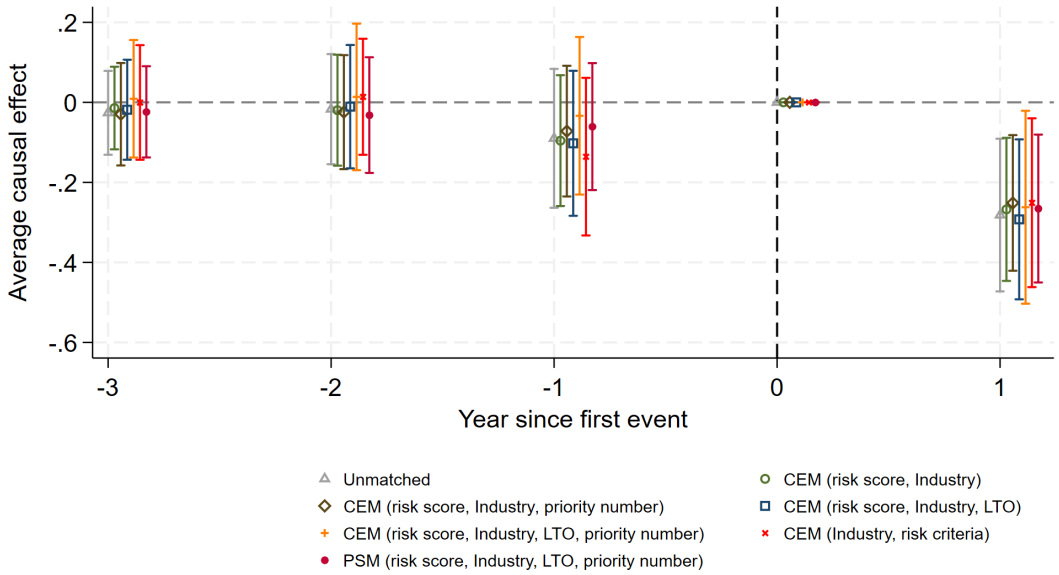
6.1 Aggregate tax liabilities

Figure 6 tests the presence of common pre-treatment trends across treatment cohorts by fitting Equation 7. The results suggest the absence of statistically significant differences in trends between audited and matched control units before audits. Post-treatment coefficients instead provide a first attempt to estimate the impact of audits on the change in SEs' reportings, evidencing a counter-deterrence effect the year after the audits.

Table 1 displays the ATT estimates derived by fitting equation (6) using various sets of matching variables (Sets I-V), one for each column. The rows of the table show the results of regressions without and with control variables.

As we can see in Table 1, audited taxpayers exhibit a notable decrease (averaging approximately -17.7 log points across models with controls) in reported tax liabilities in

Figure 6: The Impact of Audits on Reported Tax Liabilities



Note: This figure reports the estimates of the period-specific treatment effects of closed tax audits obtained by fitting Equation 7. Estimates are obtained through alternative matching techniques and are based on weighted fixed effects models with time fixed effects at year×ISIC industry×LTO level. The excluded category is the last period before treatment (T=0); 95 percent confidence intervals are shown and based on S.E. clustered at LTO level.

Table 1: Main Results – Aggregate *ATT* (closed cases)

Dependent Variable	Tax liability reported				
Matching estimator	(1)	(2)	(3)	(4)	(5)
CEM	-0.157***	-0.154***	-0.173***	-0.188***	-0.171***
	(0.051)	(0.049)	(0.058)	(0.065)	(0.058)
CEM with controls	-0.177***	-0.165***	-0.176***	-0.188***	-0.179***
	(0.059)	(0.050)	(0.062)	(0.066)	(0.063)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

the year following the audit process. This equates to an average reduction of 16.2 percent in tax liability reported by audited firms compared to their matched unaudited counterparts, which in levels corresponds to a reduction of about €8,770 (see also Table 14). The results are stable across the various methods employed, showing a relatively low level of variance. To the best of our knowledge, the possibility of tax audits backfiring immediately after the audit is a new result.[28] To rationalize these results, we explore various sources of heterogeneity in the observed effect in the following subsections.

6.2 Heterogeneity in SEs' characteristics

This section investigates potential divergences in the estimated ATT across a spectrum of firm characteristics. In the interest of space, we report here and in Section 6.3 solely the results from models including control variables. Results from baseline models without controls are reported in Appendix C and corroborate this analysis.

6.2.1 Noncompliance status

In line with previous contributions in the literature (see, for example, Kleven et al., 2011; Gemmell and Ratto, 2012; Beer et al., 2020), a key focus has been on investigating the distinct impact of audits based on the results of the audit process. This involves estimating the ATT separately for audited taxpayers determined to be noncompliant and compliant. Based on this distinction, the results presented in Table 2 show that the aggregate counter deterrence effect is entirely driven by audited taxpayers who are found to be noncompliant, with no significant influence on compliant taxpayers. This result is robust across the different matching specifications, and is in line with most of the existing research that finds the overall impact to be largely shaped by noncompliant taxpayers (Gemmell and Ratto, 2012; Løyland and Øvrum, 2017; DeBacker et al., 2018*a*, and Beer et al., 2020). However, the key distinction between our findings and previous studies is that we observe a negative impact of audits on compliance, indicating that the audits may actually reduce even more the level of compliance among noncompliant taxpayers.

[28] Other contributions already evidenced the possibility of counter deterrence in post-audit reaction, but that was either limited to an audit category—as in Erard, Kirchler and Olsen (2019) and Kotsogiannis et al. (2024)—or dependent on the effectiveness of audits and on the taxpayer's prior reporting behavior, as in Kasper and Alm (2022). Linked to this, and using a survey experiment, Lancee, Rossel and Kasper (2023) provides evidence that unfairness undermines pro-deterrence effects of audits.

Table 2: *ATT* by Audit Outcome - Noncompliant *versus* Compliant (with controls)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.076	-0.065	-0.063	-0.101	-0.100
	(0.065)	(0.057)	(0.068)	(0.078)	(0.067)
Noncompliant	-0.275***	-0.269***	-0.317***	-0.241***	-0.221***
	(0.072)	(0.070)	(0.073)	(0.081)	(0.082)
Linear combinations					
Audit+Noncompliant	-0.350***	-0.334***	-0.380***	-0.342***	-0.321***
	(0.072)	(0.065)	(0.074)	(0.076)	(0.086)
Observations	4,939	4,829	4,122	3,180	3812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

6.2.2 Magnitude of noncompliance and tax burden

The question that naturally arises is whether along the uncovered noncompliance distribution, the backlash of audits varies. One might expect, for example, that firms where significant noncompliance has been discovered respond with less reporting of the liabilities the year after, in order to recover part of the loss incurred through the audit. Table 3 illustrates precisely this, showing that with the exception of the second quartile, as the level of noncompliance detected increases, taxpayers tend to declare less income in the year following the audit.

One might also expect that the tax burden of a firm matters for noncompliance, as discussed in Slemrod (2007) and Slemrod and Gillitzer (2014). To investigate this, we estimate the audits' impact by quartile of the Effective Tax Rate (ETR), defined as the ratio between tax liabilities and taxable income in the year before the audit. Table 4 presents the results, revealing that the impact is significantly more pronounced in higher quartiles. This implies that the counter deterrence effect intensifies with the ETR, indicating a heightened impact for firms experiencing a greater tax pressure.

Finally, existing empirical evidence seems to indicate that, although there is substantial noncompliance among small businesses (Slemrod and Gillitzer, 2014), the noncompliance

Table 3: *ATT* by Quartiles of Noncompliance (with controls)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.188**	-0.195***	-0.201**	-0.211**	-0.171**
	(0.081)	(0.072)	(0.094)	(0.099)	(0.081)
Audit×Quartile 2	0.029	0.050	-0.042	-0.045	-0.019
	(0.127)	(0.125)	(0.158)	(0.164)	(0.123)
Audit×Quartile 3	-0.200**	-0.174*	-0.230**	-0.214*	-0.201*
	(0.093)	(0.095)	(0.095)	(0.121)	(0.118)
Audit×Quartile 4	-0.529**	-0.480**	-0.504**	-0.318	-0.595**
	(0.207)	(0.210)	(0.240)	(0.255)	(0.228)
Linear combinations					
Audit + Audit×Quartile 2	-0.159	-0.144	-0.243**	-0.256**	-0.190*
	(0.098)	(0.092)	(0.116)	(0.128)	(0.109)
Audit + Audit×Quartile 3	-0.387***	-0.369***	-0.431***	-0.426***	-0.372***
	(0.110)	(0.108)	(0.113)	(0.120)	(0.120)
Audit + Audit×Quartile 4	-0.716***	-0.674***	-0.705***	-0.529**	-0.766***
	(0.180)	(0.186)	(0.201)	(0.218)	(0.221)
Observations	3,666	3,573	3,024	2,303	2,733

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

rate is lower as firm size increases Slemrod (2007). Given our focus here is on SEs, we want to explore whether there are any significant differences in their responses across SEs of different size. Larger SEs may have more established procedures in place and more external oversight that make them less immune to meddling with their tax declarations. Table 5 shows that the impact of the tax audit on noncompliance does not seem statistically distinguishable across quartiles of declared income. However, the precision in the point estimate is notably higher for medium-sized (second quartile) and large SEs (fourth quartile). This suggests that it is the tax pressure, in conjunction with the outcome of the audit, rather than the size of the SE, that plays a pivotal role in shaping the compliance response. What could explain these results? Intuition might suggest that the factors contributing to the counter deterrence effect of tax auditing are intertwined with loss re-

Table 4: *ATT* by Effective Tax Rate (ETR) Quartile (with controls)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.106	-0.164*	-0.161	-0.230**	-0.066
	(0.106)	(0.098)	(0.100)	(0.115)	(0.106)
Audit×Quartile 2	-0.364***	-0.423***	-0.390***	-0.488***	-0.312***
	(0.087)	(0.066)	(0.105)	(0.116)	(0.088)
Audit×Quartile 3	-0.568***	-0.636***	-0.652***	-0.796***	-0.592***
	(0.076)	(0.064)	(0.101)	(0.117)	(0.082)
Audit×Quartile 4	-1.178***	-1.279***	-1.225***	-1.437***	-1.216***
	(0.097)	(0.102)	(0.135)	(0.206)	(0.131)
Linear combinations					
Audit + Audit×Quartile 2	-0.470***	-0.587***	-0.550***	-0.718***	-0.378**
	(0.169)	(0.138)	(0.178)	(0.198)	(0.182)
Audit + Audit×Quartile 3	-0.674***	-0.800***	-0.813***	-1.026***	-0.658***
	(0.154)	(0.130)	(0.162)	(0.197)	(0.166)
Audit + Audit×Quartile 4	-1.284***	-1.443***	-1.386***	-1.667***	-1.282***
	(0.172)	(0.163)	(0.199)	(0.271)	(0.209)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

pair motivations (as explored by Andreoni, Erard and Feinstein (1998) and Maciejovsky, Kirchler and Schwarzenberger (2007)), misperceptions of the likelihood of a future audit resulting in a bomb-crater effect (akin to Mittone (2006)), and the potential inaccuracies within the audit process, as outlined and discussed in Section 3. Moreover, all these aspects might be further exacerbated during, and soon after, an economic crisis.

6.2.3 Industry

Another factor that may help explain our results is the potential variation in responses across different types of business activities. Sectors, broadly classified as services or manufacturing, differ in their customer-facing nature, giving them differential opportunities for cash transactions and hence tax evasion. In Table 6 we estimate the differential impact of audits on reported tax liabilities across industry classification (first digit, ISIC).

Table 5: *ATT* by Income Quartile (with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.209	-0.158	-0.152	-0.056	-0.330**
	(0.132)	(0.122)	(0.151)	(0.190)	(0.132)
Audit×Quartile 2	-0.047	-0.090	-0.038	-0.213	0.128
	(0.148)	(0.144)	(0.174)	(0.201)	(0.159)
Audit×Quartile 3	0.144	0.065	0.023	-0.160	0.227*
	(0.155)	(0.149)	(0.170)	(0.193)	(0.130)
Audit×Quartile 4	0.012	-0.009	-0.093	-0.129	0.193
	(0.143)	(0.142)	(0.159)	(0.219)	(0.153)
Linear combinations					
Audit + Audit×Quartile 2	-0.256***	-0.249**	-0.190*	-0.269**	-0.201*
	(0.092)	(0.094)	(0.099)	(0.115)	(0.113)
Audit + Audit×Quartile 3	-0.065	-0.093	-0.129	-0.216**	-0.103
	(0.078)	(0.071)	(0.089)	(0.088)	(0.079)
Audit + Audit×Quartile 4	-0.198**	-0.168**	-0.245***	-0.185*	-0.136*
	(0.078)	(0.073)	(0.069)	(0.099)	(0.071)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III,Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

No particular sector seems to stand out, as the interaction terms are generally not statistically different from zero. At first glance, this might seem surprising, especially since some sectors have been shown to be more prone to tax under-reporting (Artavanis, Morse and Tsoutsoura, 2016). However, the backlash against audits from noncompliant SEs that we document here appears distinct from the typical tax evasion behavior discussed in the literature, as it is not driven by the same underlying factors. Furthermore, there seem to be deficiencies in the audit process that affect firms regardless of size or industry—an issue we turn our focus to next.

Table 6: *ATT* by ISIC Industry (with controls)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.388***	-0.362***	-0.371***	-0.326**	-0.220
	(0.110)	(0.133)	(0.124)	(0.139)	(0.236)
Audit×Industry B	0.265	0.117	0.360	0.177	0.070
	(0.234)	(0.263)	(0.244)	(0.291)	(0.448)
Audit×Industry D	0.276*	0.258	0.225	0.163	0.024
	(0.164)	(0.177)	(0.180)	(0.196)	(0.241)
Audit×Industry E	0.229	0.212	0.247	0.162	0.089
	(0.150)	(0.174)	(0.180)	(0.203)	(0.287)
Audit×Industry F	0.268*	0.188	0.150	-0.060	0.031
	(0.138)	(0.145)	(0.167)	(0.220)	(0.267)
Audit×Industry G	0.321	0.346	0.388*	0.404	0.197
	(0.229)	(0.267)	(0.232)	(0.287)	(0.340)
Audit×Industry H	0.165	0.095	0.058	-0.291	-0.078
	(0.202)	(0.215)	(0.240)	(0.315)	(0.300)
Audit×Industry I	0.430	0.463	0.393	0.399	-0.107
	(0.329)	(0.328)	(0.374)	(0.398)	(0.352)
Linear combinations					
Audit + Audit×Industry B	-0.124	-0.245	-0.011	-0.150	-0.150
	(0.185)	(0.201)	(0.184)	(0.242)	(0.347)
Audit + Audit×Industry D	-0.112	-0.105	-0.146	-0.163	-0.196**
	(0.095)	(0.086)	(0.105)	(0.107)	(0.090)
Audit + Audit×Industry E	-0.159*	-0.150*	-0.124	-0.164	-0.131
	(0.082)	(0.086)	(0.111)	(0.133)	(0.113)
Audit + Audit×Industry F	-0.120	-0.174*	-0.221*	-0.386**	-0.189*
	(0.099)	(0.093)	(0.114)	(0.176)	(0.111)
Audit + Audit×Industry G	-0.067	-0.016	0.017	0.078	-0.023
	(0.181)	(0.202)	(0.201)	(0.221)	(0.200)
Audit + Audit×Industry H	-0.223	-0.267*	-0.313*	-0.617**	-0.298*
	(0.149)	(0.149)	(0.180)	(0.247)	(0.171)
Audit + Audit×Industry I	0.042	0.100	0.022	0.073	-0.327
	(0.296)	(0.316)	(0.332)	(0.366)	(0.263)
Observations	4,710	4,607	3,943	3,087	3,707

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III,Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2. Industries are defined is Section 5.1, Figure 2.

6.3 Heterogeneity in audit process

6.3.1 External *versus* internal audits triggers

There are several potential triggers for a tax audit. Following guidance from the Revenue Authority, we have grouped them in two categories: "internal" investigations originating based on the Revenue Authority's selection criteria and "external". As already noted, an audit can be initiated either based on intelligence internal to the Revenue Authority or mandated by other government bodies. Externally ordered cases arise when requests to audit are made by the Financial Crime Investigation Directorate, the Public Prosecutor's Office or another relevant judicial authority. Internal sources of information predominantly take the form of risk assessment and prioritization, but also cross-checks initiated by information and intelligence within the Revenue Authority.

As previously discussed (Section 5.1), internal audits make up the vast majority of cases—88.6 percent compared to 11.4 percent for external audits (Figure 3). However, internal audits are less effective at identifying noncompliance, flagging only 33.4 percent of cases, whereas external audits identify noncompliance in 65 percent of cases (Figure 4). It is therefore interesting to investigate if this difference in their "detection power" also arises in terms of deterrence impact. Table 7 presents the results of this. As can be seen from the table, the negative effect on compliance is solely attributable to internal audits. The counter deterrence effect is entirely and significantly neutralized when audits are externally mandated, as indicated by the interaction term, resulting in a non-significant net effect for this audit category. In other words, externally ordered audits do not significantly alter the SEs' behavior. Given that both audits are conducted by the same tax auditors, the neutral result for externally mandated audits most likely indicate inefficiency or noisiness in selecting those firms for audit rather than relying on the internally determined criteria.

6.3.2 Desk ("partial") versus on-the-spot comprehensive ("full") audits

The tax administration may carry out either a desk tax audit from its offices ("partial") or a comprehensive and on-the-spot audit ("full"). While both types of audits require the taxpayer to be notified in writing, they are substantially different enforcement instruments. Desk audits are conducted remotely and they are based on financial statements, declarations and other documents submitted by the taxpayer, as well as documents and information that the Revenue Authority has already in its possession. Full audits, instead,

Table 7: *ATT* by Type of Audit: External *versus* Internal (with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.195***	-0.179***	-0.196***	-0.215***	-0.199***
	(0.060)	(0.052)	(0.065)	(0.072)	(0.065)
Audit×External	0.149*	0.119	0.156*	0.184**	0.146*
	(0.078)	(0.076)	(0.087)	(0.089)	(0.079)
Linear combinations					
Audit + Audit×External	-0.045	-0.059	-0.040	-0.031	-0.052
	(0.077)	(0.072)	(0.081)	(0.073)	(0.073)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

entail a visit of the tax auditor to the premises of the firm and involve the inspection of a wider set of documents, including the accounting books. Therefore, they are more extensive in their scope, but also more expensive in terms of both time and manpower. As suggested by the descriptive evidence presented in Section 5.1 (Figure 4), the different amount of resources devoted to these audits is reflected in their detection capacity, with full audits being able to flag noncompliance in the 62.4 percent of cases while this figure is reduced by half for desk audits. The question we want to investigate here is whether the extra cost is justified in generating a differential response by firms in terms of future deterrence.

Table 8 presents the results of this analysis. As can be seen from this table, there is no statistical difference between partial and full audits, which is really surprising, given the additional effort and time that complete audits often demand. In contrast, Kotsogiannis et al. (2024) show that while "full" tax audits have a significant positive effect on compliance, "partial" tax audits exhibit both a positive and a negative effect during a three-year period after the audit, with the net impact being negative. Our findings here seem to indicate that firms are more concerned with the incidence of an audit than with its depth (and to some extent length).

Table 8: *ATT* by Type of audit: Partial *versus* Full (with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.170***	-0.161***	-0.162**	-0.192***	-0.162**
	(0.059)	(0.050)	(0.061)	(0.068)	(0.064)
Audit×Full	-0.035	-0.018	-0.073	0.023	-0.097
	(0.099)	(0.099)	(0.147)	(0.150)	(0.106)
Linear combinations					
Audit + Audit×Full	-0.204*	-0.179*	-0.235	-0.170	-0.259**
	(0.105)	(0.103)	(0.147)	(0.147)	(0.111)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

6.3.3 Audited period

The Revenue Authority determines the proportion of audits to perform in a given wave in relation to the inspected periods of liabilities reported. For audits carried out from January 1, 2018 (our sample) at least 75 percent of the audits by law should relate to fiscal years within the last three years and thus the financial years 2013-2017. The remaining 25 percent of audits can be from fiscal years outside this three year window. The Revenue Authority can go beyond the statute of limitation (5 years), and audit firms for their tax declarations prior to 2013 under certain conditions.[29] The question on which light is shed on here is whether the tax period for which a taxpayer is audited in a given audit wave (2018, in our case) makes any difference in terms of deterrence. To put it differently, the question an answer is sought for is whether an inspection on an old tax declaration—relative to the year of the audit—triggers a different impact with respect to an audit targeting a recently presented tax return. To the best of our knowledge, although this is an issue that has attracted a lot of discussion (see for example Mihu, 2011), it has received no empirical attention until now.

We explore this issue from two different angles. We first classify the audits into three

[29]Due to the selection procedure implemented for our sample, that entails concentrating on firms and excluding non operating businesses, the relative shares of these groups do not reflect those described above (Figure 3).

categories based on the period of examined tax declarations relative to the year of the audit. We then analyze the differential effect of audits that involve tax returns that are subject to the statute of limitation.[30] The statute of limitations limits the audits to review only the tax returns that were submitted within five years before the audit begins. However, the Revenue Authority can audit older than five year cases, but they are constrained by law as to how many of those they can audit in a given year.

As already discussed in Section 5.1, the data allows us to categorize the audit cases based on the period of examined tax declarations relative to the 2018 audit wave as follows: for cases up to three years old (denoted T_3); for cases up to five years old (denoted T_5); and for cases that received an exception to the statute of limitations on valid grounds (denoted $T_{>5}$). In Table 9 we examine the audit's impact based on these categories. While the estimated coefficients exhibit no statistical differences among the three categories, the precision of the estimation varies across cases. Notably, the category of oldest cases stands out as the only one revealing a statistically significant impact.

Table 10 presents the results of the analysis that estimate the differential impact of audits involving tax returns nearing the statute of limitation. This means that, in this case, the focus is on the cluster of audits on tax returns up to five years old (T_5) and, in particular, within this cluster on those about to expire, corresponding to tax liabilities filed in 2013. Although, on the whole, they seem statistically indistinguishable, Columns (3) and (4) stand out, revealing a stronger and statistically relevant impact leading to a stronger counter deterrence.

6.3.4 Audit intensity

The analysis now turns to the impact of audit intensity on compliance, aiming to determine whether a firm audited multiple times during the same audit wave exhibits different compliance behavior compared to one audited only once. The scant evidence on this issue points to the fact that small firms that are audited infrequently have significant noncompliance rates (Slemrod and Gillitzer, 2014). This difference in the intensity of the tax audit can lead to a different reaction because a taxpayer audited multiple times during the same

[30] Additionally, Appendix C (Tables C.8 and C.9), examines the effect of distinguishing the audit cases that review the most recent tax return from the rest. In the broader context of most models, audits linked to the most recent tax return reported reveal impacts indistinguishable from their counterparts. However, Columns (3) and (4) stand as exceptions, manifesting a distinct inclination toward a significant reversal in the sign of the impact, even though the estimated effect does not reach statistical significance.

Table 9: ATT by Audited Period: T_3, T_5, $T_{>5}$ (with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.103	-0.084	-0.114	-0.206**	-0.087
	(0.081)	(0.076)	(0.092)	(0.091)	(0.095)
Audit×T_5	0.000	-0.049	-0.061	0.110	-0.143
	(0.168)	(0.164)	(0.149)	(0.175)	(0.133)
Audit×$T_{>5}$	-0.136	-0.143	-0.098	0.024	-0.153
	(0.103)	(0.101)	(0.131)	(0.128)	(0.118)
Linear combinations					
Audit + Audit×T_5	-0.103	-0.133	-0.175	-0.097	-0.230*
	(0.159)	(0.146)	(0.130)	(0.152)	(0.117)
Audit + Audit×$T_{>5}$	-0.239***	-0.227***	-0.212**	-0.183*	-0.241***
	(0.077)	(0.071)	(0.094)	(0.105)	(0.081)
Observations	4,939	4,829	4,122	3,180	3,812

Note: ATT for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table 10: ATT by Audited Period: Statute of Limitation (with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.129	-0.261	-0.441***	-0.361**	-0.295*
	(0.273)	(0.202)	(0.159)	(0.159)	(0.156)
Audit×2013	0.046	-0.098	-0.407**	-0.353**	-0.075
	(0.232)	(0.182)	(0.179)	(0.174)	(0.119)
Linear combinations					
Audit + Audit×2013	-0.084	-0.359	-0.848***	-0.714**	-0.370
	(0.486)	(0.359)	(0.290)	(0.272)	(0.227)
Observations	706	700	606	472	604

Note: ATT for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

wave may perceive the process differently with respect to those audited once.

Table 11 highlights that the counter deterrence effect of audits is stronger for firms

audited two or more times in the same wave with respect to the others. To some extent, and in light of the results so far, the fact that multiple-audited SEs do not comply more in the year after the audit is not surprising. If audits are inefficient in the sense that they fail to uncover the extent of noncompliance, and even though they occur frequently, it may still pay off for SEs, on average, to increase their noncompliance. This, one may argue, relates to the literature investigating the impact of the length of the punishment—exposure (the "dose", in terms of the number of audits) to punishment—on offenders re-offending (the extent of tax compliance in the year after the final audit wave), with exposure here referring to the administrative costs—broadly defined to include unpleasantness and possible stigma—of the audits.[31] That might not be appropriate to conduct many intensive audits across several periods is also recognized in technical guidance of international organizations, Biber (2010).

Table 11: *ATT* by Audit Intensity: Single *versus* Multiple audited (with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.175***	-0.174***	-0.173***	-0.205***	-0.176***
	(0.054)	(0.046)	(0.060)	(0.075)	(0.066)
Multiple audits	-0.587***	-0.683***	-0.581**	-0.963*	-0.487**
	(0.183)	(0.204)	(0.263)	(0.522)	(0.185)
Linear combinations					
Audit + Multiple audits	-0.762***	-0.857***	-0.754***	-1.167**	-0.663***
	(0.204)	(0.221)	(0.280)	(0.560)	(0.220)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

6.3.5 Audit status: Closed *versus* ongoing cases

Audits take time to complete, and some audits may not finish before the firm submits its next income tax declaration. As information matters in shaping the beliefs of the taxpayer, the question then is how does the status of the audit impact behavior? Firms whose audit has been completed have all the necessary information before submitting their next income

[31] As discussed in Rose and Shem-Tov (2021) and the literature reviewed there.

tax declaration. Are SEs under investigation behaving in a similar manner or not? This is an important question, as this comparison would allow us to separate whether behavior changes based simply on the information that the firm is under investigation or whether it is the actual result of the audit that matters for compliance. While the large majority of audits in our sample are closed (2,072), there is a small cluster in the data of ongoing audits (70) that can be suitably used to make this comparison.

Table 12 displays the aggregate ATT estimated for this subset of ongoing audits, revealing a noteworthy pro-deterrence impact on reported liabilities one year after the audit process. The same result is obtained by considering the whole set of audits—including closed and ongoing cases—and by interacting the dummy identifying the ongoing cases with the audit treatment. Table 13 presents the results of this exercise. We find that ongoing audits show an effect significantly different from closed cases, leading to an aggregate impact of opposite sign. This result lead to the conclusion that the determining factor for the estimated negative impact on closed cases is the information concerning the outcome of the audit process, rather than simply being under inspection. Indeed, recall that is the noncompliant cluster that is driving the results on the aggregate for the closed cases.

Table 12: Main Results – Aggregate ATT (ongoing cases)

Dependent Variable	Tax liability reported				
Matching estimator	(1)	(2)	(3)	(4)	(5)
Ongoing audit	0.409**	0.395*	0.501*	0.567*	0.307
	(0.195)	(0.205)	(0.246)	(0.285)	(0.222)
Ongoing audit with controls	0.479**	0.469**	0.479*	0.591*	0.170
	(0.190)	(0.209)	(0.240)	(0.311)	(0.228)
Observations	2,723	1,764	463	167	755

Note: ATT for ongoing & notified audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

6.4 Returns to tax auditing

Do audits generate more revenue per euro spent on tax enforcement? This is the question addressed in this section. Tables 14-16 present a back-of-the-envelope calculation of aver-

Table 13: *ATT* by Type of Audit: Closed *versus* Ongoing Cases (with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.179***	-0.166***	-0.178***	-0.190***	-0.180***
	(0.058)	(0.050)	(0.061)	(0.065)	(0.063)
Audit×Ongoing	0.635***	0.640***	0.723***	0.834***	0.493**
	(0.203)	(0.200)	(0.225)	(0.227)	(0.230)
Linear combinations					
Audit + Audit×Ongoing	0.456**	0.474**	0.545**	0.644***	0.313
	(0.199)	(0.195)	(0.217)	(0.224)	(0.229)
Observations	5,006	4,898	4,187	3,237	3,892

Note: *ATT* for all audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

age benefits and costs associated with tax auditing for closed cases, noncompliant-closed cases and ongoing cases, respectively. The revenue impact encompasses both the revenue recovered through the detection of underreported tax liability (plus any penalties incurred) and the revenue gained through deterring noncompliance the first year after the audit, corresponding to the estimated *ATT*. The cost of tax enforcement is estimated using the salary of an auditor multiplied by the average duration of an audit within the respective category expressed relative to the average number of audits per auditor performed in a fiscal year.[32] More precisely, the cost of an audit is calculated as the average gross cost (salaries plus insurance) in 2019 multiplied by the average duration of the audit (over the year), and does not include the cost of management, as well as any overhead costs such as building and technology service costs.[33] These tables provide information on audit performance in terms of detection power as a percentage of the potential tax liability—defined as the sum of self-assessment and underreporting assessed.[34] On the aggregate, closed cases detect underreporting equal to almost 7 percent of the potential tax liability (Table 14, second row). This figure reaches 18 percent if one considers just closed audits detecting noncompliance

[32]This is based on the activity log which records the duration of audits.

[33]The return to enforcement can, therefore, be taken as an upper bound.

[34]Ideally, this figure should be calculated as a percentage of the true tax liability (see, for example, Johns and Slemrod, 2010; Guyton et al., 2021) which would also account for an estimate of undetected underreporting in the denominator. Such an estimate is however not available and therefore the detection power of audits is likely to be overestimated.

Table 14: Audits Descriptive Statistics - ALL Closed Cases

	N	Mean	SD	Min	Max
Detection (underreporting plus fines, €1,000)	2,072	13.341	93.438	0.000	1,975.251
Detection (% Potential Tax Liability)	2,072	6.724	15.878	0	100
Deterrence effect (€1,000)	2,072	-8.770	15.690	-172.392	0.000
Aggregate impact (detection plus deterrence, €1,000)	2,072	4.572	89.601	-169.792	1,893.792
Aggregate impact (% Potential Tax Liability)	2,072	-9.709	18.692	-17.700	100
Average duration (months)	2072	15.092	0	15.092	15.092
Cost of one audit (€1,000)	2,072	3.773	0.000	3.773	3.773
Benefit (net of costs, €1,000)	2,072	0.799	89.601	-173.565	1,890.019
Return-on-investment (€)	2,072	0.212	23.748	-46.002	500.932

(Table 15, second row) and to almost 36 percent if consideration is taken of the temporary assessment of ongoing cases determined by the tax auditor but not yet communicated to the taxpayers (Table 16, second row).

To estimate the return-on-investment (ROI), we calculate the ratio between the net benefit of audits and their respective costs (last row of each table), to find the net yield for every euro invested is €0.21 in the aggregate for closed cases (Table 14). The limited and ineffective impact on the aggregate is due to the importance of the counter deterrence effect relative to the limited average detection power that is somewhat more balanced for the cases detecting noncompliance. In this case, the ROI is instead around €2.28 for every euro invested (Table 15).[35] As already discussed, the deterrence effect is negative for closed cases and positive for ongoing ones. Regarding the average duration of audits, ongoing cases have started significantly earlier (relative to the audit year) than the closed ones.[36] Of note is that the (provisional) ROI for ongoing cases is €18.23 out of any euro invested (Table 16).

[35]This is in line with the aggregate estimate of US$2.21 reported for the U.S. in Boning et al. (2024), but this is driven by the noncompliant SEs.

[36]This might be the case because they are more complex audits and/or these audits have been allocated to less experienced tax auditors. Unfortunately, this is something that we cannot identify in the data.

Table 15: Audits Descriptive Statistics - Closed Cases noncompliant

	N	Mean	SD	Min	Max
Detection (underreporting + fines, €1,000)	766	36.088	151.042	0.006	1,975.251
Detection (% Potential Tax Liability)	766	18.187	21.765	0.005	100
Deterrence effect (€1,000)	766	-21.792	38.570	-336.409	0.000
Aggregate impact (detection + deterrence, €1,000)	766	14.296	140.102	-333.809	1,893.792
Aggregate impact (% Potential Tax Liability)	766	-10.071	29.283	-34.534	100
Average duration (months)	766	17.443	0	17.443	17.443
Cost of one audit (€1,000)	766	4.361	0.000	4.361	4.361
Benefit (net of costs, €1,000)	766	9.936	140.102	-338.169	1,889.432
Return-on-investment (€)	766	2.278	32.128	-77.550	433.289

Table 16: Audits Descriptive Statistics - ALL Ongoing Cases

	N	Mean	SD	Min	Max
Detection (underreporting + fines, €1,000)	70	153.494	707.441	0.100	5467.473
Detection (% Potential Tax Liability)	70	35.942	32.439	0.323	100
Deterrence effect (€1,000)	70	27.117	38.895	0.000	222.357
Aggregate impact (detection + deterrence, €1,000)	70	180.611	705.999	0.586	5467.473
Aggregate impact (% Potential Tax Liability)	70	67.1	16.66	48.806	100
Average duration (months)	70	37.564	0	37.564	37.564
Cost of one audit (€1,000)	70	9.391	0.000	9.391	9.391
Benefit (net of costs, €1,000)	70	171.220	705.999	-8.805	5458.082
Return to Investment (€)	70	18.232	75.178	-0.938	581.204

7 Concluding remarks

The safeguarding of tax compliance requires not only sufficient resources to the administration, but also a compliance strategy which verifies that the true income has been reported by the taxpayer. Tax audits are crucial for ensuring compliance with tax laws and maintaining the integrity of the tax system, ensuring that all businesses contribute their fair share to public finances. They are also important for leveling the playing field, minimizing distortions arising from noncompliance behavior.

The focus of this paper has been on whether operational audits foster compliance among small enterprises that are both populous and tend to be harder to monitor. The results suggest that audits can backfire in terms of compliance on the aggregate and across a broad range of heterogeneity related to small enterprises but also, and most importantly, across many dimensions of the audit process. Audits predominantly have a negative impact on future compliance across most margins, with one notable exception: when small enterprises file their tax return while the tax audit is already underway, but they have not yet been informed about the outcome. Despite the negative impact of audits on compliance, the marginal impact of the enforcement is on average positive, and fairly sizeable.

The analysis above is of course subject to limitations. It abstracts, for instance, from considerations of the exact reason for underreporting and in particular whether it is caused intentionally or unintentionally, and, if the latter, whether it is driven by inattention about the intricate tax system (as recently explored by Christiansen (2024) for the self-employed). It also sets aside the role of tax auditors in the audit process, or the role of the supply chain in the diffusion of audits. In recognition of the challenges posed by the SEs, who are often considered both hard-to-tax and hard-to-audit due to the nature of their income streams and record-keeping practices, the IAPR recently implemented significant reforms whose aim are to increase transparency and compliance while reducing opportunities for tax evasion. Measures include stricter reporting requirements, the compulsory usage of point-of-sales terminals and Business-to-Business digital invoicing systems. By modernizing its tax infrastructure and targeting loopholes that previously allowed underreporting, the IAPR seeks to enhance revenue collection while addressing long-standing inefficiencies in SEs compliance processes. Evaluating those reforms, and their impact on compliance, is an avenue for future research.

While these issues deserve attention, the analysis here has pointed to two key policy

lessons for Revenue Authorities. One conclusion to draw from all this is the need for more effective and careful evaluation of tax audits. There is no doubt that tax audits evaluation is now being given more importance by Revenue Authorities, yet this is done without as yet the full recognition for the need to do this to the same standards increasingly expected in other areas of public policy. To do so will require building the tax audits evaluation more purposively into project design. A second lesson is that there are important heterogeneities in the audit process, which should be part of any design considerations for an audit strategy. In retrospect, this should not come as a great surprise, yet it has not received much attention in the literature.

References

Abadie, Alberto. 2005. "Semiparametric Difference-in-Differences Estimators." *Review of Economic Studies*, 72(1): 1–19.

Advani, Arun. 2022. "Who Does and Doesn't Pay Taxes?" *Fiscal Studies*, 43(1): 5–22.

Advani, Arun, William Elming, and Jonathan Shaw. 2021. "The Dynamic Effects of Tax Audits." *The Review of Economics and Statistics*, 0(0): 1–45.

Alm, James, and Antoine Malézieux. 2021. "40 Years of Tax Evasion Games: A Meta-Analysis." *Experimental Economics,*, 0(0): 1–52.

Andreoni, James, Brian Erard, and Jonathan Feinstein. 1998. "Tax Compliance." *Journal of Economic Literature*, 36(2): 818–860.

Artavanis, Nikolaos, Adair Morse, and Margarita Tsoutsoura. 2016. "Measuring Income Tax Evasion Using Bank Credit: Evidence from Greece ." *The Quarterly Journal of Economics*, 131(2): 739–798.

Bachas, Pierre, Roberto N. Fattal Jaef, and Anders Jensen. 2019. "Size-dependent Tax Enforcement and Compliance: Global Evidence and Aggregate Implications." *Journal of Development Economics*, 140: 203–222.

Bayer, Ralph, and Frank Cowell. 2009. "Tax Compliance and Firms' Strategic Interdependence." *Journal of Public Economics*, 93(11): 1131–1143.

Beer, Sebastian, Matthias Kasper, Erich Kirchler, and Brian Erard. 2020. "Do Audits Aeter Future Noncompliance? Evidence On Self-Employed Taxpayers." *CesIfo Economic Studies,*, 66(3): 248—-264.

Best, Michael, Jawad Shah, and Mazhar Waseem. 2021. "The Deterrence Value of Tax Audit: Estimates from a Randomised Audit Program." Mimeo.

Biber, Edmund. 2010. "Revenue Administration: Taxpayer Audit—Development of Effective." IMF, Technical Guidance.

Boning, William C., Nathaniel Hendren, Ben Sprung-Keyser, and Ellen Stuart. 2024. "A Welfare Analysis of Tax Audits Across the Income Distribution." *The Quarterly Journal of Economics*, qjae037.

Brockmeyer, Anne, and Magaly Sáenz Somarriba. 2025. "Electronic Payment Technology and Tax Compliance: Evidence from Uruguay's Financial Inclusion Reform." *American Economic Journal: Economic Policy*, 17(1): 242–72.

Cai, Hongbin, and Qiao Liu. 2009. "Competition and Corporate Tax Avoidance: Evidence from Chinese Industrial Firms." *The Economic Journal*, 119(537): 764–795.

Cattaneo, Matias D. 2010. "Efficient Semiparametric Estimation of Multi-Valued Treatment Effects Under Ignorability." *Journal of Econometrics*, 155(2): 138–154.

Christiansen, Tobias Gabel. 2024. "Dynamic Effects of Tax Audits and the Role of Intentions." *Journal of Public Economics*, 234: 105121.

DeBacker, Jason, Bradley T. Heim, Anh Tran, and Alexander Yuskavage. 2018*a*. "Once Bitten, Twice Shy? The Lasting Impact of Enforcement on Tax Compliance." *Journal of Law and Economics,*, 61(1): 1–35.

DeBacker, Jason, Bradley T. Heim, Anh Tran, and Alexander Yuskavage. 2018*b*. "The Effects of IRS Audits on EITC Claimants." *National Tax Journal,*, 71(3): 451–484.

Erard, Brian, Erich Kirchler, and Jerome Olsen. 2019. "Audit Impact Study: The Specific Deterrence Implications of Increased Reliance on Correspondence Audits." 257–268.

Gemmell, Norman, and Marisa Ratto. 2012. "Behavioral Responses to Taxpayer Audits: Evidence From Random Taxpayer Inquiries." *National Tax Journal,*, 65(1): 33–57.

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

Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman. 2021. "Taxation of SMEs in OECD and G20 Countries." NBER Working Paper 28542.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies,*, 64(4): 605–654.

Heckman, James J., Hidehiko Ichimura, and Petra Todd. 1998. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies,*, 65(2): 261–294.

Henning, David, Christos Kotsogiannis, Luca Salvadori, and Jukka Pirttilä. 2024. "Audits and Tax Compliance: Evidence from Uganda." Mimeo.

HMRC. 2024. "Measuring Tax Gaps 2024 Edition: Tax Gap Estimates for 2022 to 2023." HMRC Official Statistics.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2011. "Multivariate Matching Methods That Are Monotonic Imbalance Bounding." *Journal of the American Statistical Association,*, 106(493): 345–361.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2012. "Causal Inference Without Balance Checking: Coarsened Exact Matching." *Political Analysis,*, 20(1): 1–24.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2019. "A Theory of Statistical Inference for Matching Methods in Causal Research." *Political Analysis*, 27(1): 46–68.

ILO. 2019. "Small Matters. Global Evidence on the Contribution to Employment by the Self-employed, micro-enterprises and SMEs." International Labour Organization - ILO publishing.

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

Johns, Andrew, and Joel Slemrod. 2010. "The Distribution of Income Tax Noncompliance." *National Tax Journal*, 63(3): 397–418.

Kasper, Matthias, and James Alm. 2022. "Audits, Audit Effectiveness, and Post-Audit Tax Compliance." *Journal of Economic Behavior & Organization*, 195: 87–102.

Keen, Michael, and Joel Slemrod. 2017. "Optimal Tax Administration." *Journal of Public Economics*, 152: 133–142.

King, Gary, Richard Nielsen, Carter Coberley, James E. Pope, and Aaron Wells. 2011. "Comparative Effectiveness of Matching Methods for Causal Inference." Mimeo.

Kirchler, Erich, Erik Hoelzl, and Ingrid Wahl. 2008. "Enforced versus Voluntary Tax Compliance: The "Slippery Slope" Framework." *Journal of Economic Psychology*, 29(2): 210–225.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. 2011. "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica,*, 79(3): 651–692.

Kotsogiannis, Christos, Luca Salvadori, John Karangwa, and Innocente Murasi. 2025. "E-invoicing, Tax Audits and VAT Compliance." *Journal of Development Economics*, 172: 103403.

Kotsogiannis, Christos, Luca Salvadori, John Karangwa, and Theonille Mukamana. 2024. "Do Tax Audits Have a Dynamic Impact? Evidence from Corporate Income Tax Administrative Data." *Journal of Development Economics*, 170: 103292.

Lancee, Bora, Lucia Rossel, and Matthias Kasper. 2023. "When the Agency Wants Too Much: Experimental Evidence on Unfair Audits and Tax Compliance." *Journal of Economic Behavior & Organization*, 214: 406–442.

Løyland, Knut, and Arnstein Øvrum. 2017. "The Effect of Targeted Audits on Immigrant Workers' Tax Compliance." Working Paper.

Maciejovsky, Boris, Erich Kirchler, and Herbert Schwarzenberger. 2007. "Misperception of Chance and Loss Repair: On the Dynamics of Tax Compliance." *Journal of Economic Psychology*, 28(6): 678–691.

Mihu, George. 2011. "The Tax Audit – Procedural Aspects and Consequences." *Economics, Management, and Financial Markets*, 6: 934–941.

Mittal, Shekhar, and Aprajit Mahajan. 2017. "Vat in Emerging Economies: Does Third Party Verification Matter?" mimeo.

Mittone, Luigi. 2006. "Dynamic Behaviour in Tax Evasion: An Experimental Approach." *The Journal of Socio-Economics*, 35(5): 813–835.

Naritomi, Joana. 2019. "Consumers as Tax Auditors." *American Economic Review*, 109(9): 3031–72.

OECD. 2015. "Taxation of SMEs in OECD and G20 Countries." OECD Tax Policy Studies, No. 23, OECD Publishing, Paris.

Pomeranz, Dina, and José Vila-Belda. 2019. "Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities." *Annual Review of Economics*, 11(1): 755–781.

Rose, Evan K., and Yotam Shem-Tov. 2021. "How Does Incarceration Affect Reoffending? Estimating the Dose-Response Function." *Journal of Political Economy*, 129(12): 3302–3356.

Slemrod, Joel. 2007. "Cheating Ourselves: The Economics of Tax Evasion." *Journal of Economic Perspectives*, 21(1): 25–48.

Slemrod, Joel. 2019. "Tax Compliance and Enforcement." *Journal of Economic Literature,*, 57(4): 904–954.

Slemrod, Joel, and Christian Gillitzer. 2014. "Editor's Choice Insights from a Tax-systems Perspective." *CESifo Economic Studies,*, 60(1): 1–31.

Slemrod, Joel, Brett Collins, Jeffrey L. Hoopes, Daniel Reck, and Michael Sebastiani. 2017. "Does Credit-Card Information Reporting Improve Small-Business Tax Compliance?" *Journal of Public Economics*, 149(C): 1–19.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian. 2001. "Taxpayer Response to An Increased Probability of Audit: Evidence From a Controlled Experiment in Minnesota." *Journal of Public Economics*, 79(3): 455–483.

Stuart, Elizabeth A. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science,*, 25(1): 1–21.

Wahl, Ingrid, Barbara Kastlunger, and Erich Kirchler. 2010. "Trust in Authorities and Power to Enforce Tax Compliance: An Empirical Analysis of the "Slippery Slope Framework"." *Law & Policy*, 32(4): 383–406.

Wang, Yiqun. 2012. "Competition and Tax Evasion: A Cross Country Study." *Economic Analysis and Policy*, 42(2): 189–208.

Wooldridge, Jeffrey M. 2002. "Inverse Probability Weighted M-estimators for Sample Selection, Attrition, and Stratification." *Portuguese Economic Journal*, 1(2): 117–139.

Wooldridge, Jeffrey M. 2007. "Inverse Probability Weighted Estimation for General Missing Data Problems." *Journal of Econometrics*, 141(2): 1281–1301.

Online Appendices

Appendix A: Outcomes of the CEM procedure

Several matching approaches are discussed in the existing literature, with the goal of seeking a trade-off between improving balance among relevant pre-treatment covariates and maintaining a reasonable sample size for matching (see, among others, King et al., 2011; Imbens and Rubin, 2015; Guo and Fraser, 2015). Our preference is Coarsened Exact Matching (CEM) due to its robust statistical properties, surpassing common methods like Propensity Score Matching and Mahalanobis Distance Matching in mitigating initial imbalances.[1]

Additionally, the CEM algorithm is highly intuitive. It categorizes pre-treatment variables into meaningful bins based on user-defined thresholds or standard binning algorithms.[2] Units with the same bin signature are placed in a stratum, and control units within each stratum are weighted to match the number of treated units. Strata without both treated and control units are removed. Using these weights in equation (5), we employ a Difference-in-Differences (DID) approach to estimate the ATT of audits.

Specifically, CEM is applied to stratify the sample based on variables described in Section 5.2, and the results of the matching procedures are presented in Tables A.1-A.5. These tables include a summary of matched and unmatched observations (Panel A), a measure of multivariate imbalance reduction (Panel B), and univariate imbalance statistics for each matching variable, including the difference at the mean and other relevant points of the covariates' distribution across treatment cohorts (Panel C). The imbalance measures are based on the L_1 difference between the multivariate or univariate histogram of matching variables across treatment cohorts (see Iacus, King and Porro, 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.

Across all tables, these measures show a substantial reduction in imbalance after CEM,

[1] CEM also minimizes model dependence, estimation error, bias, variance, and other criteria while optimizing the trade-off between sample size and balance (see Iacus, King and Porro, 2011, 2012; Blackwell et al., 2009; King et al., 2011; King and Nielsen, 2019 for more details and formal proofs, and Iacus, King and Porro, 2019 for a discussion on the inference theory).

[2] Specifically, for categorical variables (ISIC industry and LTO), we perform exact matching, and for other variables, we use deciles.

Table A.1: Summary of the CEM Matching Procedure: First Set of Covariates

Panel A: Matching summary

	Non-audited	Audited
Available observations	2,990	2,072
Matched	2,973	2,072
Unmatched	17	0

Panel B: Overall imbalance, Multivariate L_1

L_1 statistic pre CEM:	0.330
L_1 statistic post CEM:	0.255

Panel C: Difference across treatment cohorts

Before CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.278	104.486	-3.2	8.67	108.5	607.5	-1855
Industry	0.152	0.111	0	0	1	0	0
After CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	.220	33.487	5.35	0	99.33	-104	225
Industry	4.27e-15	-1.33e-14	0	0	0	0	.

Note: The table depicts the matching summary of the CEM procedure, L_1 statistics for imbalance as defined in Iacus, King and Porro (2011), and differences across treatment cohorts in the distribution of the risk scores before and after CEM. Matching covariates are described in Section 5.2. Industries are grouped according to the high-level ISIC aggregation: A=Agriculture, forestry and fishing; B=Manufacturing, mining and quarrying and other industry; C=Construction; D=Wholesale and retail trade, transportation and storage, accommodation and food service activities; E=Information and communication; F=Financial and insurance activities; G=Real estate activities; H=Professional, scientific, technical, administration and support service activities; I=Public administration, defence, education, human health and social work activities.

indicating increased homogeneity in the pre-treatment covariates across treatment cohorts while maintaining a high number of matched observations.

Table A.2: Summary of the CEM Matching Procedure: Second Set of Covariates

Panel A: Matching summary

	Non-audited	Audited
Available observations	2,990	2,072
Matched	2,886	2,047
Unmatched	104	25

Panel B: Overall imbalance, Multivariate L_1

L_1 statistic pre CEM:	0.546
L_1 statistic post CEM:	0.442

Panel C: Difference across treatment cohorts

Before CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.278	104.486	-3.2	8.67	108.5	607.5	-1855
Industry	0.152	.1106672	0	0	1	0	0
Priority Number	0.273	-199.814	0	-126	-226	-265	-3
After CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.217	31.189	5.35	8	99	-145	-145
Industry	6.57e-15	1.15e-14	0	0	0	0	.
Priority Number	0.065	-9.243	0	-5	-6	-14	-2

Note: The table depicts the matching summary of the CEM procedure, L_1 statistics for imbalance as defined in Iacus, King and Porro (2011), and differences across treatment cohorts in the distribution of the risk scores before and after CEM. Matching covariates are described in Section 5.2. Industries are grouped according to the high-level ISIC aggregation: A=Agriculture, forestry and fishing; B=Manufacturing, mining and quarrying and other industry; C=Construction; D=Wholesale and retail trade, transportation and storage, accommodation and food service activities; E=Information and communication; F=Financial and insurance activities; G=Real estate activities; H=Professional, scientific, technical, administration and support service activities; I=Public administration, defence, education, human health and social work activities.

Table A.3: Summary of the CEM Matching Procedure: Third Set of Covariates

Panel A: Matching summary

	Non-audited	Audited
Available observations	2,990	2,072
Matched	2,483	1,726
Unmatched	507	346

Panel B: Overall imbalance, Multivariate L_1

L_1 statistic pre CEM:	0.475
L_1 statistic post CEM:	0.462

Panel C: Difference across treatment cohorts

Before CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.278	104.486	-3.2	8.67	108.5	607.5	-1855
Industry	0.152	0.111	0	0	1	0	0
Local Tax Offices	0.138	396.133	0	2	900	689	0
After CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.316	33.403	5.35	-1	87	93	0
Industry	2.78e-15	8.88e-16	0	0	0	0	.
Local Tax Offices	3.18e-15	5.46e-12	0	0	0	0	0

Note: The table depicts the matching summary of the CEM procedure, L_1 statistics for imbalance as defined in Iacus, King and Porro (2011), and differences across treatment cohorts in the distribution of the risk scores before and after CEM. Matching covariates are described in Section 5.2. Industries are grouped according to the high-level ISIC aggregation: A=Agriculture, forestry and fishing; B=Manufacturing, mining and quarrying and other industry; C=Construction; D=Wholesale and retail trade, transportation and storage, accommodation and food service activities; E=Information and communication; F=Financial and insurance activities; G=Real estate activities; H=Professional, scientific, technical, administration and support service activities; I=Public administration, defence, education, human health and social work activities.

Table A.4: Summary of the CEM Matching Procedure: Fourth Set of Covariates

Panel A: Matching summary

	Non-audited	Audited
Available observations	2,990	2,072
Matched	1,838	1,398
Unmatched	1,152	674

Panel B: Overall imbalance, Multivariate L_1

L_1 statistic pre CEM:	0.711
L_1 statistic post CEM:	0.527

Panel C: Difference across treatment cohorts

Before CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.278	104.486	-3.2	8.67	108.5	607.5	-1855
Industry	0.152	0.111	0	0	1	0	0
Local Tax Offices	0.138	396.133	0	2	900	611	0
Priority Number	0.273	-199.814	0	-126	-226	-265	-3
After CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.198	27.390	2.55	21	84	-19.5	0
Industry	1.37e-15	-8.88e-16	0	0	0	0	.
Local Tax Offices	1.91e-15	5.91e-12	0	0	-79	0	0
Priority Number	0.066	-5.812	1	-5	-8	-10	-10

Note: The table depicts the matching summary of the CEM procedure, L_1 statistics for imbalance as defined in Iacus, King and Porro (2011), and differences across treatment cohorts in the distribution of the risk scores before and after CEM. Matching covariates are described in Section 5.2. Industries are grouped according to the high-level ISIC aggregation: A=Agriculture, forestry and fishing; B=Manufacturing, mining and quarrying and other industry; C=Construction; D=Wholesale and retail trade, transportation and storage, accommodation and food service activities; E=Information and communication; F=Financial and insurance activities; G=Real estate activities; H=Professional, scientific, technical, administration and support service activities; I=Public administration, defence, education, human health and social work activities.

Table A.5: Summary of the CEM Matching Procedure: Fifth Set of Covariates

Panel A: Matching summary

	Non-audited	Audited
Available observations	2,990	2,072
Matched	2,148	1,735
Unmatched	842	335

Panel B: Overall imbalance, Multivariate L_1

L_1 statistic pre CEM:	0.497
L_1 statistic post CEM:	0.018

Panel C: Difference across treatment cohorts

Before CEM							
	L1	mean	min	25%	50%	75%	max
Industry	0.152	0.111	0	0	1	0	0
Risk criteria (average)	0.039	-0.098	-0.615	0.000	-1.346	0.000	-1.231
After CEM							
	L1	mean	min	25%	50%	75%	max
Industry	2.90e-15	-1.15e-14	0	0	0	0	.
Risk criteria (average)	0.000	-0.005	0.000	0.000	-0.179	0.000	0.000

Note: The table depicts the matching summary of the CEM procedure, L_1 statistics for imbalance as defined in Iacus, King and Porro (2011), and differences across treatment cohorts in the distribution of the risk scores before and after CEM. Matching covariates are described in Section 5.2. Industries are grouped according to the high-level ISIC aggregation: A=Agriculture, forestry and fishing; B=Manufacturing, mining and quarrying and other industry; C=Construction; D=Wholesale and retail trade, transportation and storage, accommodation and food service activities; E=Information and communication; F=Financial and insurance activities; G=Real estate activities; H=Professional, scientific, technical, administration and support service activities; I=Public administration, defence, education, human health and social work activities.

Appendix B: Outcomes of the CEM procedure for ongoing cases

Table B.1: Summary of the CEM Matching Procedure: First Set of Covariates - Ongoing Cases

Panel A: Matching summary

	Non-audited	Audited
Available observations	2,990	70
Matched	2,712	68
Unmatched	278	2

Panel B: Overall imbalance, Multivariate L_1

L_1 statistic pre CEM:	0.463
L_1 statistic post CEM:	0.343

Panel C: Difference across treatment cohorts

Before CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.293	537.446	67.2	16.17	69.5	887.5	-1765
Industry	0.163	-0.328	0	0	0	0	0
After CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.179	13.846	82	6	23	-15	-30
Industry	1.63e-14	1.07e-14	0	0	0	0	.

Note: The table depicts the matching summary of the CEM procedure, L_1 statistics for imbalance as defined in Iacus, King and Porro (2011), and differences across treatment cohorts in the distribution of the risk scores before and after CEM. Matching covariates are described in Section 5.2. Industries are grouped according to the high-level ISIC aggregation: A=Agriculture, forestry and fishing; B=Manufacturing, mining and quarrying and other industry; C=Construction; D=Wholesale and retail trade, transportation and storage, accommodation and food service activities; E=Information and communication; F=Financial and insurance activities; G=Real estate activities; H=Professional, scientific, technical, administration and support service activities; I=Public administration, defence, education, human health and social work activities.

Table B.2: Summary of the CEM Matching Procedure: Second Set of Covariates - Ongoing Cases

Panel A: Matching summary

	Non-audited	Audited
Available observations	2,990	70
Matched	1,730	66
Unmatched	1,260	4

Panel B: Overall imbalance, Multivariate L_1

L_1 statistic pre CEM:	0.798
L_1 statistic post CEM:	0.568

Panel C: Difference across treatment cohorts

Before CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.293	537.446	67.2	16.17	69.5	887.5	-1765
Industry	0.163	-0.328	0	0	0	0	0
Priority Number	0.413	-250.904	0	-179	-316	-404	-1074
After CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.186	11.606	76.5	11	4	46.67	-30
Industry	1.38e-15	1.07e-14	0	0	0	0	.
Priority Number	0.117	-10.798	0	-14	-27	-11	-98

Note: The table depicts the matching summary of the CEM procedure, L_1 statistics for imbalance as defined in Iacus, King and Porro (2011), and differences across treatment cohorts in the distribution of the risk scores before and after CEM. Matching covariates are described in Section 5.2. Industries are grouped according to the high-level ISIC aggregation: A=Agriculture, forestry and fishing; B=Manufacturing, mining and quarrying and other industry; C=Construction; D=Wholesale and retail trade, transportation and storage, accommodation and food service activities; E=Information and communication; F=Financial and insurance activities; G=Real estate activities; H=Professional, scientific, technical, administration and support service activities; I=Public administration, defence, education, human health and social work activities.

Table B.3: Summary of the CEM Matching Procedure: Third Set of Covariates - Ongoing Cases

Panel A: Matching summary

	Non-audited	Audited
Available observations	2,990	70
Matched	422	51
Unmatched	2,568	19

Panel B: Overall imbalance, Multivariate L_1

L_1 statistic pre CEM:	0.758
L_1 statistic post CEM:	0.474

Panel C: Difference across treatment cohorts

Before CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.293	537.446	67.2	16.17	69.5	887.5	-1765
Industry	0.163	-0.328	0	0	0	0	0
Local Tax Offices	0.312	522.799	0	2	1781	478	0
After CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	.170	17.680	37.5	6	2	390.83	-163.33
Industry	5.33e-16	-2.66e-15	0	0	0	0	.
Local Tax Offices	5.74e-16	-4.55e-12	0	0	0	0	0

Note: The table depicts the matching summary of the CEM procedure, L_1 statistics for imbalance as defined in Iacus, King and Porro (2011), and differences across treatment cohorts in the distribution of the risk scores before and after CEM. Matching covariates are described in Section 5.2. Industries are grouped according to the high-level ISIC aggregation: A=Agriculture, forestry and fishing; B=Manufacturing, mining and quarrying and other industry; C=Construction; D=Wholesale and retail trade, transportation and storage, accommodation and food service activities; E=Information and communication; F=Financial and insurance activities; G=Real estate activities; H=Professional, scientific, technical, administration and support service activities; I=Public administration, defence, education, human health and social work activities.

Table B.4: Summary of the CEM Matching Procedure: Forth Set of Covariates - Ongoing Cases

Panel A: Matching summary

	Non-audited	Audited
Available observations	2,990	70
Matched	132	38
Unmatched	2,858	32

Panel B: Overall imbalance, Multivariate L_1

L_1 statistic pre CEM:	0.932
L_1 statistic post CEM:	0.182

Panel C: Difference across treatment cohorts

Before CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.293	537.446	67.2	16.17	69.5	887.5	-1765
Industry	0.163	-0.328	0	0	0	0	0
Local Tax Offices	0.312	522.799	0	2	1781	478	0
Priority Number	0.413	-250.904	0	-179	-316	-404	-1074
After CEM							
	L1	mean	min	25%	50%	75%	max
Risk Score	0.079	36.738	10.5	17	10.5	145	-163.33
Industry	1.24e-15	1.78e-15	0	0	0	0	0
Local Tax Offices	1.02e-15	1.64e-11	0	0	0	0	0
Priority Number	0.031	-2.666	-1	-7	-34	-9	-49

Note: The table depicts the matching summary of the CEM procedure, L_1 statistics for imbalance as defined in Iacus, King and Porro (2011), and differences across treatment cohorts in the distribution of the risk scores before and after CEM. Matching covariates are described in Section 5.2. Industries are grouped according to the high-level ISIC aggregation: A=Agriculture, forestry and fishing; B=Manufacturing, mining and quarrying and other industry; C=Construction; D=Wholesale and retail trade, transportation and storage, accommodation and food service activities; E=Information and communication; F=Financial and insurance activities; G=Real estate activities; H=Professional, scientific, technical, administration and support service activities; I=Public administration, defence, education, human health and social work activities.

Table B.5: Summary of the CEM Matching Procedure: Fifth Set of Covariates - Ongoing Cases

Panel A: Matching summary

	Non-audited	Audited
Available observations	2,990	70
Matched	716	51
Unmatched	2,274	19

Panel B: Overall imbalance, Multivariate L_1

L_1 statistic pre CEM:	0.812
L_1 statistic post CEM:	0.044

Panel C: Difference across treatment cohorts

Before CEM							
	L1	mean	min	25%	50%	75%	max
Industry	0.163	-0.328	0	0	0	0	0
Risk criteria (average)	0.073	0.069	9.000	0.000	-1.885	-0.077	-4.077
After CEM							
	L1	mean	min	25%	50%	75%	max
Industry	1.56e-15	-1.78E-15	0	0	0	0	.
Risk criteria (average)	0.003	0.006	0.000	0.000	0.000	0.000	0.000

Note: The table depicts the matching summary of the CEM procedure, L_1 statistics for imbalance as defined in Iacus, King and Porro (2011), and differences across treatment cohorts in the distribution of the risk scores before and after CEM. Matching covariates are described in Section 5.2. Industries are grouped according to the high-level ISIC aggregation: A=Agriculture, forestry and fishing; B=Manufacturing, mining and quarrying and other industry; C=Construction; D=Wholesale and retail trade, transportation and storage, accommodation and food service activities; E=Information and communication; F=Financial and insurance activities; G=Real estate activities; H=Professional, scientific, technical, administration and support service activities; I=Public administration, defence, education, human health and social work activities.

Appendix C: Additional results

Table C.1: *ATT* by Audit Outcome - Noncompliant *versus* Compliant (baseline)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.054	-0.052	-0.061	-0.097	-0.080
	(0.059)	(0.057)	(0.065)	(0.075)	(0.062)
Noncompliant	-0.278***	-0.272***	-0.322***	-0.250***	-0.248***
	(0.071)	(0.069)	(0.072)	(0.079)	(0.079)
Linear combinations					
Audit+Noncompliant	-0.332***	-0.324***	-0.383***	-0.348***	-0.328***
	(0.065)	(0.062)	(0.071)	(0.076)	(0.080)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.2: *ATT* by Quartiles of Noncompliance (baseline)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.144*	-0.149**	-0.179*	-0.186*	-0.132*
	(0.075)	(0.072)	(0.092)	(0.101)	(0.072)
Audit×Quartile 2	0.007	0.027	-0.043	-0.051	-0.041
	(0.125)	(0.124)	(0.156)	(0.165)	(0.117)
Audit×Quartile 3	-0.229**	-0.219**	-0.260***	-0.239**	-0.235**
	(0.091)	(0.090)	(0.091)	(0.111)	(0.109)
Audit×Quartile 4	-0.537**	-0.514**	-0.560**	-0.416	-0.618***
	(0.202)	(0.206)	(0.236)	(0.253)	(0.223)
Linear combinations					
Audit + Audit×Quartile 2	-0.138	-0.122	-0.223**	-0.237*	-0.172*
	(0.090)	(0.087)	(0.111)	(0.126)	(0.102)
Audit + Audit×Quartile 3	-0.373***	-0.368***	-0.439***	-0.425***	-0.367***
	(0.104)	(0.101)	(0.107)	(0.108)	(0.108)
Audit + Audit×Quartile 4	-0.681***	-0.663***	-0.740***	-0.602***	-0.749***
	(0.175)	(0.182)	(0.197)	(0.215)	(0.217)
Observations	3,666	3,573	3,024	2,303	2,733

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.3: *ATT* by Effective Tax Rate (ETR) Quartile (baseline)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.115	-0.155	-0.173	-0.244**	-0.066
	(0.104)	(0.095)	(0.105)	(0.121)	(0.101)
Audit×Quartile 2	-0.356***	-0.397***	-0.379***	-0.453***	-0.311***
	(0.083)	(0.062)	(0.102)	(0.117)	(0.085)
Audit×Quartile 3	-0.575***	-0.627***	-0.643***	-0.774***	-0.599***
	(0.074)	(0.062)	(0.099)	(0.121)	(0.081)
Audit×Quartile 4	-1.171***	-1.264***	-1.217***	-1.418***	-1.213***
	(0.092)	(0.096)	(0.132)	(0.201)	(0.127)
Linear combinations					
Audit + Audit×Quartile 2	-0.470***	-0.552***	-0.552***	-0.697***	-0.378**
	(0.163)	(0.131)	(0.178)	(0.200)	(0.174)
Audit + Audit×Quartile 3	-0.689***	-0.782***	-0.815***	-1.019***	-0.665***
	(0.151)	(0.125)	(0.163)	(0.201)	(0.162)
Audit + Audit×Quartile 4	-1.285***	-1.419***	-1.390***	-1.663***	-1.279***
	(0.165)	(0.153)	(0.200)	(0.269)	(0.201)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table C.4: *ATT* by Income Quartile (baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.184	-0.138	-0.141	-0.021	-0.310**
	(0.128)	(0.124)	(0.150)	(0.197)	(0.130)
Audit×Quartile 2	-0.056	-0.105	-0.050	-0.248	0.112
	(0.149)	(0.147)	(0.180)	(0.211)	(0.162)
Audit×Quartile 3	0.122	0.047	-0.003	-0.207	0.193
	(0.154)	(0.150)	(0.167)	(0.194)	(0.130)
Audit×Quartile 4	0.004	-0.015	-0.104	-0.178	0.193
	(0.139)	(0.140)	(0.157)	(0.227)	(0.150)
Linear combinations					
Audit + Audit×Quartile 2	-0.240***	-0.243**	-0.191*	-0.269**	-0.199*
	(0.089)	(0.093)	(0.097)	(0.110)	(0.112)
Audit + Audit×Quartile 3	-0.062	-0.091	-0.144	-0.228**	-0.117
	(0.075)	(0.070)	(0.087)	(0.087)	(0.075)
Audit + Audit×Quartile 4	-0.180***	-0.153**	-0.245***	-0.199**	-0.117*
	(0.067)	(0.068)	(0.064)	(0.096)	(0.065)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III,Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.5: *ATT* by ISIC Industry (baseline)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.402***	-0.369***	-0.363***	-0.316**	-0.231
	(0.116)	(0.134)	(0.123)	(0.137)	(0.241)
Audit×Industry B	0.271	0.097	0.352	0.160	0.084
	(0.237)	(0.263)	(0.242)	(0.286)	(0.432)
Audit×Industry D	0.315*	0.278	0.219	0.149	0.052
	(0.163)	(0.177)	(0.178)	(0.193)	(0.244)
Audit×Industry E	0.262*	0.237	0.240	0.151	0.105
	(0.151)	(0.178)	(0.179)	(0.201)	(0.288)
Audit×Industry F	0.255*	0.163	0.151	-0.059	0.006
	(0.146)	(0.152)	(0.163)	(0.215)	(0.276)
Audit×Industry G	0.354	0.357	0.371	0.382	0.189
	(0.233)	(0.268)	(0.232)	(0.284)	(0.343)
Audit×Industry H	0.188	0.108	0.044	-0.301	-0.017
	(0.204)	(0.215)	(0.239)	(0.310)	(0.301)
Linear combinations					
Audit + Audit×Industry B	-0.131	-0.272	-0.012	-0.156	-0.147
	(0.186)	(0.203)	(0.182)	(0.241)	(0.345)
Audit + Audit×Industry D	-0.086	-0.091	-0.145	-0.168	-0.179**
	(0.086)	(0.084)	(0.104)	(0.106)	(0.086)
Audit + Audit×Industry E	-0.140*	-0.131	-0.123	-0.165	-0.126
	(0.081)	(0.087)	(0.111)	(0.132)	(0.108)
Audit + Audit×Industry F	-0.146	-0.206**	-0.212*	-0.375**	-0.225**
	(0.099)	(0.094)	(0.108)	(0.169)	(0.110)
Audit + Audit×Industry G	-0.048	-0.012	0.008	0.066	-0.042
	(0.174)	(0.200)	(0.196)	(0.222)	(0.199)
Audit + Audit×Industry H	-0.214	-0.261*	-0.319*	-0.617**	-0.248
	(0.148)	(0.148)	(0.178)	(0.244)	(0.169)
Audit + Audit×Industry I	0.031	0.052	0.011	0.040	-0.402
	(0.299)	(0.326)	(0.331)	(0.370)	(0.271)
Observations	4,710	4,607	3,943	3,087	3,707

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III,Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Industries are defined is Section 5.1, Figure 2.

Table C.6: *ATT* by Type of Audit: External *versus* Internal (baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.177***	-0.172***	-0.198***	-0.221***	-0.191***
	(0.051)	(0.049)	(0.059)	(0.067)	(0.058)
Audit×External	0.165**	0.155**	0.185**	0.231***	0.154**
	(0.072)	(0.071)	(0.079)	(0.072)	(0.069)
Linear combinations					
Audit + Audit×External	-0.011	-0.016	-0.013	0.010	-0.037
	(0.071)	(0.072)	(0.080)	(0.074)	(0.071)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.7: *ATT* by Type of Audit: Partial *versus* Full (baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.148***	-0.146***	-0.153**	-0.176**	-0.144**
	(0.053)	(0.049)	(0.058)	(0.068)	(0.059)
Audit×Full	-0.048	-0.039	-0.118	-0.064	-0.137
	(0.105)	(0.103)	(0.144)	(0.152)	(0.108)
Linear combinations					
Audit + Audit×Full	-0.196*	-0.185*	-0.271*	-0.240	-0.281**
	(0.102)	(0.103)	(0.142)	(0.145)	(0.107)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.8: *ATT* by Audited Period: Previous Fiscal Year (baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.153***	-0.144***	-0.161***	-0.167**	-0.168***
	(0.050)	(0.049)	(0.058)	(0.066)	(0.060)
Audit×2017	0.023	0.114	0.324***	0.250**	0.095
	(0.197)	(0.140)	(0.098)	(0.096)	(0.082)
Linear combinations					
Audit + Audit×2017	-0.130	-0.030	0.163	0.083	-0.073
	(0.198)	(0.147)	(0.116)	(0.140)	(0.113)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.9: *ATT* by Audited Period: Previous Fiscal Year (with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.169***	-0.147***	-0.156**	-0.166**	-0.173**
	(0.058)	(0.050)	(0.063)	(0.068)	(0.066)
Audit×2017	-0.012	0.067	0.254**	0.135	0.055
	(0.214)	(0.154)	(0.101)	(0.096)	(0.083)
Linear combinations					
Audit + Audit×2017	-0.181	-0.080	0.098	-0.031	-0.118
	(0.211)	(0.157)	(0.119)	(0.135)	(0.108)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table C.10: ATT by Audited Period: T_3, T_5, $T_{>5}$ (baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.092	-0.083	-0.113	-0.218**	-0.097
	(0.078)	(0.073)	(0.088)	(0.090)	(0.088)
Audit×T_5	0.008	-0.016	-0.042	0.104	-0.117
	(0.167)	(0.167)	(0.146)	(0.171)	(0.123)
Audit×$T_{>5}$	-0.131	-0.132	-0.121	0.042	-0.116
	(0.094)	(0.094)	(0.117)	(0.125)	(0.106)
Linear combinations					
Audit + Audit×T_5	-0.084	-0.099	-0.155	-0.114	-0.214*
	(0.155)	(0.149)	(0.126)	(0.147)	(0.108)
Audit + Audit×$T_{>5}$	-0.223***	-0.215***	-0.234***	-0.176*	-0.213***
	(0.067)	(0.066)	(0.085)	(0.102)	(0.070)
Observations 4,939	4,829	4,122	3,180	3,812	680

Note: ATT for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.11: ATT by Audited Period: Statute of Limitation (baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.094	-0.212	-0.424***	-0.382***	-0.206*
	(0.255)	(0.186)	(0.140)	(0.141)	(0.121)
Audit×2013	-0.046	-0.199	-0.473**	-0.565**	-0.073
	(0.254)	(0.238)	(0.180)	(0.233)	(0.138)
Linear combinations					
Audit + Audit×2013	-0.140	-0.411	-0.897***	-0.946***	-0.279
	(0.474)	(0.370)	(0.266)	(0.296)	(0.212)
Observations	706	700	606	472	604

Note: ATT for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.12: *ATT* by Audit Intensity: Single *versus* Multiple Audited (baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.158***	-0.165***	-0.172***	-0.205***	-0.169***
	(0.046)	(0.044)	(0.055)	(0.073)	(0.060)
Multiple audits	-0.540***	-0.664***	-0.581**	-0.904*	-0.445**
	(0.174)	(0.195)	(0.252)	(0.477)	(0.174)
Linear combinations					
Audit + Multiple audits	-0.698***	-0.829***	-0.753***	-1.109**	-0.614***
	(0.188)	(0.210)	(0.268)	(0.515)	(0.200)
Observations	4,939	4,829	4,122	3,180	3,812

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.13: *ATT* by Type of Audit: Closed *versus* Ongoing Cases (baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.157***	-0.154***	-0.173***	-0.188***	-0.170***
	(0.051)	(0.049)	(0.057)	(0.064)	(0.057)
Audit×Ongoing	0.564***	0.558***	0.648***	0.761***	0.454**
	(0.202)	(0.198)	(0.229)	(0.233)	(0.221)
Linear combinations					
Audit + Audit×Ongoing	0.406**	0.404**	0.475**	0.573**	0.283
	(0.194)	(0.190)	(0.220)	(0.227)	(0.215)
Observations	5,006	4,898	4,187	3,237	3,892

Note: *ATT* for all audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix D: Robustness analysis: Propensity score matching

Table D.1: Main Results (PSM) – Aggregate *ATT*

Dependent Variable	Tax liability reported				
Matching estimator	(1)	(2)	(3)	(4)	(5)
PSM	-0.153***	-0.154***	-0.168***	-0.176***	-0.175***
	(0.053)	(0.053)	(0.055)	(0.056)	(0.056)
PSM with controls	-0.184***	-0.175***	-0.178***	-0.174***	-0.195***
	(0.059)	(0.055)	(0.057)	(0.056)	(0.061)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III,Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table D.2: *ATT* by Audit Outcome - Noncompliant *versus* Compliant (PSM - baseline)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.045	-0.046	-0.060	-0.067	-0.067
	(0.059)	(0.060)	(0.060)	(0.061)	(0.063)
Noncompliant	-0.291***	-0.291***	-0.291***	-0.291***	-0.289***
	(0.071)	(0.071)	(0.071)	(0.071)	(0.071)
Linear combinations					
Audit+Noncompliant	-0.336***	-0.337***	-0.351***	-0.358***	-0.357***
	(0.068)	(0.069)	(0.070)	(0.071)	(0.069)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.3: *ATT* by Audit Outcome - Noncompliant *versus* Compliant (PSM with controls)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.077	-0.067	-0.066	-0.066	-0.090
	(0.063)	(0.060)	(0.060)	(0.061)	(0.067)
Noncompliant	-0.287***	-0.289***	-0.292***	-0.292***	-0.289***
	(0.071)	(0.071)	(0.070)	(0.070)	(0.071)
Linear combinations					
Audit+Noncompliant	-0.364***	-0.356***	-0.359***	-0.358***	-0.378***
	(0.074)	(0.070)	(0.074)	(0.072)	(0.074)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table D.4: *ATT* by Quartiles of Noncompliance (PSM baseline)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.151**	-0.153**	-0.167**	-0.174**	-0.171**
	(0.075)	(0.075)	(0.075)	(0.074)	(0.074)
Audit×Quartile 2	-0.016	-0.016	-0.016	-0.016	-0.019
	(0.121)	(0.121)	(0.121)	(0.121)	(0.121)
Audit×Quartile 3	-0.218**	-0.218**	-0.218**	-0.218**	-0.221**
	(0.086)	(0.086)	(0.086)	(0.086)	(0.087)
Audit×Quartile 4	-0.535**	-0.535**	-0.535**	-0.535**	-0.537**
	(0.226)	(0.226)	(0.226)	(0.226)	(0.226)
Linear combinations					
Audit + Audit×Quartile 2	-0.168*	-0.169*	-0.183**	-0.190**	-0.189**
	(0.087)	(0.087)	(0.088)	(0.089)	(0.089)
Audit + Audit×Quartile 3	-0.370***	-0.371***	-0.385***	-0.392***	-0.391***
	(0.090)	(0.090)	(0.094)	(0.093)	(0.090)
Audit + Audit×Quartile 4	-0.686***	-0.687***	-0.701***	-0.709***	-0.708***
	(0.200)	(0.201)	(0.201)	(0.203)	(0.202)
Observations	3,569	3,569	3,569	3,569	3,565

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.5: *ATT* by Quartiles of Noncompliance (PSM with controls)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.224***	-0.219***	-0.208**	-0.218***	-0.251***
	(0.082)	(0.079)	(0.080)	(0.076)	(0.080)
Audit×Quartile 2	0.027	0.030	0.040	0.041	0.034
	(0.122)	(0.120)	(0.120)	(0.118)	(0.121)
Audit×Quartile 3	-0.167*	-0.157	-0.161*	-0.153	-0.150
	(0.096)	(0.096)	(0.096)	(0.095)	(0.098)
Audit×Quartile 4	-0.500**	-0.490**	-0.498**	-0.480**	-0.491**
	(0.231)	(0.229)	(0.231)	(0.232)	(0.231)
Linear combinations					
Audit + Audit×Quartile 2	-0.197**	-0.189**	-0.168*	-0.177**	-0.217**
	(0.088)	(0.086)	(0.088)	(0.087)	(0.094)
Audit + Audit×Quartile 3	-0.392***	-0.376***	-0.370***	-0.371***	-0.402***
	(0.100)	(0.099)	(0.102)	(0.102)	(0.102)
Audit + Audit×Quartile 4	-0.724***	-0.709***	-0.706***	-0.698***	-0.743***
	(0.206)	(0.204)	(0.205)	(0.209)	(0.209)
Observations	3,569	3,569	3,569	3,569	3,565

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table D.6: *ATT* by Effective Tax Rate (ETR) Quartile (PSM - baseline)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.083	-0.082	-0.089	-0.087	-0.071
	(0.073)	(0.070)	(0.075)	(0.071)	(0.075)
Audit×Quartile 2	-0.390***	-0.378***	-0.367***	-0.381***	-0.379***
	(0.070)	(0.065)	(0.084)	(0.079)	(0.073)
Audit×Quartile 3	-0.606***	-0.614***	-0.585***	-0.587***	-0.576***
	(0.062)	(0.063)	(0.071)	(0.069)	(0.066)
Audit×Quartile 4	-1.221***	-1.260***	-1.233***	-1.217***	-1.210***
	(0.080)	(0.080)	(0.099)	(0.086)	(0.092)
Linear combinations					
Audit + Audit×Quartile 2	-0.473***	-0.460***	-0.456***	-0.468***	-0.450***
	(0.124)	(0.111)	(0.139)	(0.126)	(0.129)
Audit + Audit×Quartile 3	-0.689***	-0.696***	-0.675***	-0.674***	-0.648***
	(0.116)	(0.109)	(0.128)	(0.119)	(0.120)
Audit + Audit×Quartile 4	-1.304***	-1.342***	-1.322***	-1.305***	-1.281***
	(0.126)	(0.123)	(0.148)	(0.133)	(0.139)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.7: *ATT* by Effective Tax Rate (ETR) Quartile (PSM with controls)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit					
Audit	-0.082	-0.092	-0.082	-0.083	-0.085
	(0.074)	(0.070)	(0.074)	(0.071)	(0.077)
Audit×Quartile 2	-0.398***	-0.396***	-0.388***	-0.402***	-0.398***
	(0.070)	(0.067)	(0.085)	(0.079)	(0.073)
Audit×Quartile 3	-0.601***	-0.618***	-0.602***	-0.608***	-0.593***
	(0.063)	(0.065)	(0.073)	(0.070)	(0.065)
Audit×Quartile 4	-1.223***	-1.269***	-1.241***	-1.230***	-1.224***
	(0.084)	(0.085)	(0.102)	(0.089)	(0.094)
Linear combinations					
Audit + Audit×Quartile 2	-0.480***	-0.487***	-0.470***	-0.485***	-0.482***
	(0.122)	(0.112)	(0.136)	(0.126)	(0.130)
Audit + Audit×Quartile 3	-0.683***	-0.710***	-0.683***	-0.691***	-0.677***
	(0.114)	(0.109)	(0.127)	(0.119)	(0.118)
Audit + Audit×Quartile 4	-1.305***	-1.361***	-1.323***	-1.313***	-1.309***
	(0.132)	(0.129)	(0.149)	(0.135)	(0.143)
Observations	4786	4786	4786	4786	4780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table D.8: *ATT* by Income Quartile (PSM baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.207	-0.210	-0.223	-0.249	-0.245*
	(0.146)	(0.145)	(0.151)	(0.155)	(0.146)
Audit×Quartile 2	-0.026	-0.028	0.009	0.023	-0.022
	(0.163)	(0.160)	(0.168)	(0.167)	(0.152)
Audit×Quartile 3	0.162	0.145	0.134	0.147	0.161
	(0.158)	(0.155)	(0.158)	(0.156)	(0.153)
Audit×Quartile 4	0.044	0.076	0.041	0.093	0.103
	(0.143)	(0.143)	(0.148)	(0.155)	(0.141)
Linear combinations					
Audit + Audit×Quartile 2	-0.233**	-0.238**	-0.214**	-0.226**	-0.267***
	(0.090)	(0.095)	(0.092)	(0.096)	(0.091)
Audit + Audit×Quartile 3	-0.046	-0.064	-0.089	-0.102	-0.084
	(0.070)	(0.068)	(0.070)	(0.066)	(0.069)
Audit + Audit×Quartile 4	-0.163***	-0.134**	-0.182***	-0.156***	-0.142**
	(0.053)	(0.053)	(0.054)	(0.054)	(0.054)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.9: *ATT* by Income quartile (PSM with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.240	-0.234	-0.233	-0.250	-0.269*
	(0.152)	(0.145)	(0.152)	(0.156)	(0.150)
Audit×Quartile 2	-0.022	-0.019	0.007	0.025	-0.013
	(0.163)	(0.158)	(0.167)	(0.166)	(0.149)
Audit×Quartile 3	0.180	0.164	0.150	0.166	0.182
	(0.159)	(0.154)	(0.159)	(0.157)	(0.155)
Audit×Quartile 4	0.041	0.065	0.035	0.081	0.096
	(0.148)	(0.146)	(0.153)	(0.160)	(0.145)
Linear combinations					
Audit + Audit×Quartile 2	-0.262***	-0.252**	-0.227**	-0.225**	-0.282***
	(0.092)	(0.097)	(0.094)	(0.097)	(0.095)
Audit + Audit×Quartile 3	-0.060	-0.070	-0.084	-0.084	-0.087
	(0.072)	(0.068)	(0.072)	(0.067)	(0.071)
Audit + Audit×Quartile 4	-0.199***	-0.168***	-0.199***	-0.169***	-0.173***
	(0.061)	(0.056)	(0.058)	(0.056)	(0.059)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table D.10: *ATT* by ISIC Industry (PSM baseline)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.392***	-0.352***	-0.342***	-0.322***	-0.371***
	(0.114)	(0.118)	(0.121)	(0.118)	(0.113)
Audit × Audit×Industry B	0.236	0.110	0.127	0.016	0.041
	(0.239)	(0.263)	(0.244)	(0.254)	(0.254)
Audit × Audit×Industry D	0.254	0.212	0.188	0.156	0.186
	(0.159)	(0.161)	(0.168)	(0.159)	(0.160)
Audit × Audit×Industry E	0.235	0.179	0.183	0.160	0.188
	(0.149)	(0.155)	(0.152)	(0.156)	(0.166)
Audit × Audit×Industry F	0.264*	0.199	0.121	0.100	0.152
	(0.150)	(0.152)	(0.148)	(0.148)	(0.147)
Audit × Audit×Industry G	0.363	0.361	0.325	0.276	0.354
	(0.223)	(0.256)	(0.219)	(0.248)	(0.243)
Audit × Audit×Industry H	0.187	0.145	0.128	0.079	0.166
	(0.199)	(0.203)	(0.221)	(0.228)	(0.207)
Linear combinations					
Audit + Audit×Industry B	-0.156	-0.242	-0.215	-0.306	-0.331
	(0.188)	(0.211)	(0.195)	(0.212)	(0.203)
Audit + Audit×Industry D	-0.138	-0.140	-0.154*	-0.166*	-0.185**
	(0.087)	(0.087)	(0.090)	(0.087)	(0.089)
Audit + Audit×Industry E	-0.156*	-0.173*	-0.160*	-0.162*	-0.183*
	(0.082)	(0.087)	(0.086)	(0.092)	(0.101)
Audit + Audit×Industry F	-0.128	-0.153	-0.221**	-0.222**	-0.219**
	(0.102)	(0.105)	(0.087)	(0.095)	(0.101)
Audit + Audit×Industry G	-0.029	0.009	-0.017	-0.046	-0.017
	(0.165)	(0.202)	(0.162)	(0.200)	(0.189)
Audit + Audit×Industry H	-0.205	-0.207	-0.215	-0.243	-0.205
	(0.146)	(0.150)	(0.160)	(0.165)	(0.151)
Audit + Audit×Industry I	0.023	0.018	-0.015	-0.047	-0.002
	(0.300)	(0.315)	(0.308)	(0.320)	(0.316)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Industries are defined is Section 5.1, Figure 2.

Table D.11: *ATT* by ISIC Industry (PSM with controls)

Dependent Variable	Tax liability reported				
	(1)	(2)	(3)	(4)	(5)
Audit	-0.394***	-0.349***	-0.332***	-0.304**	-0.367***
	(0.109)	(0.118)	(0.118)	(0.118)	(0.112)
Audit × Audit×Industry B	0.232	0.114	0.155	0.034	0.053
	(0.234)	(0.257)	(0.237)	(0.248)	(0.251)
Audit × Audit×Industry D	0.212	0.178	0.151	0.130	0.149
	(0.159)	(0.161)	(0.168)	(0.161)	(0.163)
Audit × Audit×Industry E	0.203	0.141	0.155	0.129	0.157
	(0.146)	(0.153)	(0.151)	(0.157)	(0.168)
Audit × Audit×Industry F	0.272*	0.208	0.141	0.128	0.173
	(0.142)	(0.146)	(0.143)	(0.145)	(0.144)
Audit × Audit×Industry G	0.327	0.332	0.291	0.253	0.333
	(0.221)	(0.254)	(0.221)	(0.249)	(0.247)
Audit × Audit×Industry H	0.174	0.120	0.142	0.073	0.146
	(0.197)	(0.199)	(0.218)	(0.226)	(0.205)
Linear combinations					
Audit + Audit×Industry B	-0.163	-0.236	-0.178	-0.270	-0.314
	(0.185)	(0.206)	(0.188)	(0.206)	(0.201)
Audit + Audit×Industry D	-0.183*	-0.171*	-0.182*	-0.175*	-0.217**
	(0.097)	(0.091)	(0.094)	(0.089)	(0.095)
Audit + Audit×Industry E	-0.191**	-0.208**	-0.177**	-0.176*	-0.210*
	(0.082)	(0.087)	(0.084)	(0.092)	(0.105)
Audit + Audit×Industry F	-0.122	-0.142	-0.191**	-0.176*	-0.194*
	(0.105)	(0.107)	(0.089)	(0.096)	(0.103)
Audit + Audit×Industry G	-0.068	-0.018	-0.042	-0.051	-0.034
	(0.173)	(0.201)	(0.170)	(0.199)	(0.192)
Audit + Audit×Industry H	-0.220	-0.230	-0.190	-0.231	-0.220
	(0.147)	(0.147)	(0.160)	(0.162)	(0.150)
Audit + Audit×Industry I	0.015	0.025	0.006	-0.016	0.018
	(0.294)	(0.304)	(0.302)	(0.314)	(0.308)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2. Industries are defined is Section 5.1, Figure 2.

Table D.12: *ATT* by Type of Audit: Partial *versus* Full (PSM baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.136**	-0.138**	-0.152***	-0.159***	-0.158***
	(0.054)	(0.053)	(0.055)	(0.056)	(0.055)
Audit×Full	-0.082	-0.082	-0.082	-0.082	-0.081
	(0.112)	(0.112)	(0.112)	(0.112)	(0.111)
Linear combinations					
Audit + Audit×Full	-0.218*	-0.219*	-0.233**	-0.241**	-0.240**
	(0.111)	(0.112)	(0.111)	(0.114)	(0.114)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.13: *ATT* by Type of Audit: Partial *versus* Full (PSM with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.172***	-0.162***	-0.166***	-0.164***	-0.184***
	(0.058)	(0.055)	(0.056)	(0.056)	(0.060)
Audit×Full	-0.059	-0.062	-0.051	-0.050	-0.060
	(0.105)	(0.102)	(0.102)	(0.100)	(0.103)
Linear combinations					
Audit + Audit×Full	-0.230**	-0.225**	-0.218**	-0.214**	-0.244**
	(0.111)	(0.106)	(0.107)	(0.106)	(0.113)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table D.14: *ATT* by Type of Audit: External *versus* Internal (PSM baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.168***	-0.169***	-0.183***	-0.191***	-0.190***
	(0.055)	(0.055)	(0.056)	(0.058)	(0.058)
Audit×External	0.122*	0.122*	0.122*	0.122*	0.126*
	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)
Linear combinations					
Audit + Audit×External	-0.046	-0.047	-0.061	-0.069	-0.065
	(0.068)	(0.069)	(0.069)	(0.070)	(0.069)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.15: *ATT* by Type of Audit: External *versus* Internal (PSM with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.194***	-0.183***	-0.186***	-0.183***	-0.205***
	(0.062)	(0.058)	(0.059)	(0.059)	(0.064)
Audit×External	0.078	0.071	0.072	0.072	0.075
	(0.078)	(0.078)	(0.078)	(0.078)	(0.080)
Linear combinations					
Audit + Audit×External	-0.116	-0.112	-0.114	-0.111	-0.130*
	(0.075)	(0.072)	(0.075)	(0.072)	(0.074)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table D.16: *ATT* by Audited Period: Previous Fiscal Year (PSM baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.153***	-0.151**	-0.169***	-0.176***	-0.178***
	(0.057)	(0.057)	(0.058)	(0.059)	(0.059)
Audit×2017	0.122	0.210*	0.101	0.124	0.040
	(0.103)	(0.107)	(0.113)	(0.113)	(0.099)
Linear combinations					
Audit + Audit×2017	-0.031	0.059	-0.068	-0.052	-0.138
	(0.129)	(0.127)	(0.131)	(0.131)	(0.126)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.17: *ATT* by Audited Period: Previous Fiscal Year (PSM with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.178***	-0.164***	-0.170***	-0.166***	-0.189***
	(0.064)	(0.059)	(0.061)	(0.060)	(0.065)
Audit×2017	0.123	0.201*	0.098	0.109	0.049
	(0.106)	(0.113)	(0.116)	(0.120)	(0.111)
Linear combinations					
Audit + Audit×2017	-0.055	0.038	-0.072	-0.056	-0.140
	(0.134)	(0.133)	(0.134)	(0.135)	(0.134)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table D.18: ATT by Audited Period: T_3, T_5, $T_{>5}$ (PSM baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.115	-0.101	-0.133	-0.138*	-0.113
	(0.079)	(0.073)	(0.084)	(0.073)	(0.075)
Audit×T_5	-0.110	-0.148	-0.097	-0.075	-0.091
	(0.116)	(0.131)	(0.124)	(0.120)	(0.127)
Audit×$T_{>5}$	-0.079	-0.084	-0.074	-0.064	-0.105
	(0.093)	(0.091)	(0.101)	(0.090)	(0.089)
Linear combinations					
Audit + Audit×T_5	-0.225**	-0.249**	-0.229**	-0.214*	-0.203*
	(0.107)	(0.115)	(0.114)	(0.111)	(0.114)
Audit + Audit×$T_{>5}$	-0.194***	-0.184**	-0.207***	-0.202***	-0.217***
	(0.071)	(0.073)	(0.072)	(0.074)	(0.071)
Observations	4,786	4,786	4,786	4,786	4,780

Note: ATT for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.19: ATT by Audited Period: T_3, T_5, $T_{>5}$ (PSM with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.121	-0.106	-0.116	-0.123	-0.115
	(0.080)	(0.074)	(0.080)	(0.074)	(0.076)
Audit×T_5	-0.144	-0.189	-0.146	-0.118	-0.135
	(0.121)	(0.130)	(0.126)	(0.121)	(0.129)
Audit×$T_{>5}$	-0.092	-0.099	-0.084	-0.075	-0.130
	(0.102)	(0.099)	(0.102)	(0.095)	(0.099)
Linear combinations					
Audit + Audit×T_5	-0.265**	-0.295**	-0.262**	-0.241**	-0.250**
	(0.110)	(0.113)	(0.118)	(0.113)	(0.118)
Audit + Audit×$T_{>5}$	-0.213**	-0.205**	-0.200**	-0.199**	-0.245***
	(0.082)	(0.081)	(0.077)	(0.078)	(0.084)
Observations	4,786	4,786	4,786	4,786	4,780

Note: ATT for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table D.20: *ATT* by Audited Period: Statute of Limitation (PSM baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.194	-0.273**	-0.169	-0.199	-0.161
	(0.126)	(0.121)	(0.132)	(0.131)	(0.130)
Audit×2013	-0.011	-0.092	0.037	-0.034	0.010
	(0.134)	(0.153)	(0.139)	(0.132)	(0.120)
Linear combinations					
Audit + Audit×2013	-0.205	-0.366	-0.132	-0.232	-0.151
	(0.220)	(0.227)	(0.230)	(0.223)	(0.210)
Observations	683	683	683	683	682

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.21: *ATT* by Audited Period: Statute of Limitation (PSM with controls)

	(1)	(2)	(3)	(4)	(5)	(6)
Audit	-0.297**	-0.328**	-0.230	-0.231	-0.228	-0.224
	(0.142)	(0.136)	(0.144)	(0.147)	(0.152)	(0.144)
Audit×2013	-0.092	-0.104	-0.001	-0.048	-0.047	-0.000
	(0.117)	(0.132)	(0.136)	(0.134)	(0.119)	(0.129)
Linear combinations						
Audit + Audit×2013	-0.389*	-0.432*	-0.231	-0.279	-0.275	-0.225
	(0.222)	(0.227)	(0.235)	(0.242)	(0.232)	(0.234)
Observations	683	683	683	683	682	683

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

Table D.22: *ATT* by Audit Intensity: Single *versus* Multiple Audited (PSM baseline)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.160***	-0.167***	-0.171***	-0.190***	-0.183***
	(0.046)	(0.048)	(0.049)	(0.051)	(0.049)
Multiple audits	-0.510***	-0.653***	-0.506**	-0.711***	-0.539***
	(0.176)	(0.164)	(0.200)	(0.245)	(0.166)
Linear combinations					
Audit + Multiple audits	-0.670***	-0.819***	-0.678***	-0.901***	-0.722***
	(0.188)	(0.182)	(0.218)	(0.269)	(0.183)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.23: *ATT* by Audit Intensity: Single *versus* Multiple Audited (PSM with controls)

	(1)	(2)	(3)	(4)	(5)
Audit	-0.187***	-0.186***	-0.180***	-0.188***	-0.201***
	(0.054)	(0.050)	(0.052)	(0.052)	(0.055)
Multiple audits	-0.584***	-0.733***	-0.588***	-0.805***	-0.616***
	(0.192)	(0.179)	(0.214)	(0.264)	(0.185)
Linear combinations					
Audit + Multiple audits	-0.771***	-0.920***	-0.768***	-0.993***	-0.817***
	(0.212)	(0.199)	(0.234)	(0.288)	(0.208)
Observations	4,786	4,786	4,786	4,786	4,780

Note: *ATT* for closed audit cases and positive income reported. Any column is based on a different set of matching variables, as described in Section 5.2. In particular, Column 1 corresponds to Set I, Column 2 to Set II, Column 3 to Set III, Column 4 to Set IV and Column 5 to Set V. Standard errors clustered at LTO level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Control variables are described in Section 5.2.

References

Blackwell, Matthew, Stefano M. Iacus, Gary King, and Giuseppe Porro. 2009. "Cem: Coarsened Exact Matching in Stata." *The Stata Journal,*, 9(4): 524–546.

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

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2011. "Multivariate Matching Methods That Are Monotonic Imbalance Bounding." *Journal of the American Statistical Association,*, 106(493): 345–361.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2012. "Causal Inference Without Balance Checking: Coarsened Exact Matching." *Political Analysis,*, 20(1): 1–24.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2019. "A Theory of Statistical Inference for Matching Methods in Causal Research." *Political Analysis*, 27(1): 46–68.

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

King, Gary, and Richard Nielsen. 2019. "Why Propensity Scores Should Not Be Used for Matching." *Political Analysis,*, 27(4): 435–454.

King, Gary, Richard Nielsen, Carter Coberley, James E. Pope, and Aaron Wells. 2011. "Comparative Effectiveness of Matching Methods for Causal Inference." Mimeo.