



The deterrence effect of real-world operational tax audits on self-employed taxpayers: evidence from Italy

Gabriele Mazzolini¹ · Laura Pagani²  · Alessandro Santoro²

Accepted: 29 September 2021 / Published online: 1 November 2021
© The Author(s) 2021

Abstract

We use a large administrative tax-returns panel dataset merged with a tax audit database to estimate the effect of real-world operational tax audits on subsequent tax behavior of a large sample of Italian self-employed taxpayers. Results from operational audits do not suffer from the fact that taxpayers are aware that they have been randomly selected for research purposes and, then, such audits are viewed as more of a signal about true audit rates by the taxpayer. Our empirical approach relies on fixed-effects difference-in-difference comparisons with an ex-ante matched sample of non-audited taxpayers. To address concerns about the endogenous selection into audit, we provide evidence for the common trends assumption. We find a positive and lasting effect of audits on subsequent reported income. However, in line with theoretical predictions, taxpayers do not increase tax compliance when the tax authority does not assess a positive additional income. Our results are robust to a variety of specifications and samples.

Keywords Tax compliance · Administrative panel data · Tax audits

JEL Classification H26 · C23 · C55

✉ Laura Pagani
laura.pagani@unimib.it

Gabriele Mazzolini
gabrielemaria.mazzolini@agenziaentrate.it

Alessandro Santoro
alessandro.santoro@unimib.it

¹ Italian Revenue Agency, Rome, Italy

² Department of Economics, Management and Statistics (DEMS), Università di Milano-Bicocca, Milan, Italy

1 Introduction

In the benchmark economic model of tax-compliance, utility maximizing taxpayers decide the level of income to report considering the private costs and benefits of evasion. Compliance depends on evasion detection probability and on related penalties (Allingham & Sandmo, 1972). Tax audits are the main evasion deterrence instrument. They produce two effects on tax compliance (Gemmel & Ratto, 2012). First, when taxpayers are subject to an audit, the tax authority could detect non-compliance and force compliance, raising an additional tax yield. Second, taxpayers experiencing an audit are likely to increase their perceived probability of detection (Kleven et al., 2011). In the expected utility maximizing model of tax-compliance, this probability update will lead to an increase in their future reporting.¹

In this paper, we study audits' effect on subsequent tax reporting of audited taxpayers focusing in particular on those audits conducted by revenue agencies during their customary auditing activity and which, following Slemrod (2016), we call 'real-world operational audits.' We use a large administrative tax-returns panel dataset of Italian taxpayers whose income is obtained from self-employment and from sole proprietorships merged with a tax audit database, both made available by the Italian revenue agency.

Thanks to the recent wider availability of confidential taxpayers' administrative datasets and to the willingness of some revenue authorities to conduct field experiments to evaluate the effect of audits, there is a growing body of literature that uses randomized audits to analyze the causal effects of audits on deterrence and taxpayers' behavior. Overall, this literature has found a positive and significant effect of audits on subsequent tax compliance, although the magnitude is both heterogeneous across income types and time-variant. Kleven et al. (2011) conducted a tax-enforcement field experiment in Denmark and found that the overall effect of audits on total net income is positive but quite modest and driven entirely by self-reported income. DeBacker et al. (2018) used IRS data and found that on average audits cause a 0.4% increase of reported wage income over three years after the audit. The effect is much higher when considering self-employment income (7.5%). Advani et al. (2019) used random audits in the United Kingdom. They found a large and persistent impact of audits on reported tax liability that reached a remarkable 26% increase four years after the audit.

The previous studies of tax audits on individuals are based on random audits. These latter have undoubtedly the methodological advantage over real-world operational audits that derives from the randomness of treatment assignment. Then, the internal validity of these studies cannot be questioned. However, the use of random audits generally implies that taxpayers are aware that they have been randomly selected to be audited for research reasons, and such awareness may imply that they revise the audit risk differently from the case in which they have been chosen by the tax authority for an operational audit (Slemrod, 2016). This argument applies to De

¹ Audits may also produce spillovers to non-audited taxpayers.

Backer et al. (2018) and to Advani et al. (2019), as well as to other studies using random audits with the exception of Kleven et al. (2011).² The study of audits' effect using operational audits allows to determine more realistically the costs and benefits of audit policies and, overall, the policy relevance of results obtained using real-world tax audits may be larger compared to random audits.

There are few studies focusing on real-world operational audits.³ D'Agosto et al. (2018) study operational audits' effect using our same data source in the 2004–2009 period and find a positive deterrence effect of tax audits. However, they use a difference-in-difference model without taking explicitly into account individual time invariant unobserved heterogeneity, contrary to what we do in this paper using individual fixed effects. Moreover, with respect to the previous study, in our paper we enrich the analysis of audits' effect, exploring heterogeneity across different audit outcomes. Specifically, we exploit unique information available in our data on 'null-outcome' audits, namely cases when the revenue agency recognizes that it has mistakenly claimed additional taxes and cancels the preliminary adjustment, and we look at differential audits' effects for null and non-null audit outcomes.

An additional study using operational audits is Løyland et al. (2019), who analyze compliance effect of risk-based tax audit in Norway. Differently from our study, however, they exclude self-employed taxpayers and use self-reported deductions among wage earners and transfer recipients as outcome. They find a positive effect of audits on future compliance in terms of a fall in self-reported deductions.

Our paper analyses the impact of real-world operational audits on high-risk taxpayers. The outcome we consider is total personal income from self-employment and sole proprietorship as reported on taxpayers' tax returns. Taxes on this type of income are clearly more subject to evasion through misreporting or underreporting income compared to wage income as they are not subject to third-party reporting.

We use a fixed-effects difference-in-difference comparison with matched audited and non-audited taxpayers to address endogeneity of operational audits related to time-invariant factors. To address potential treatment assignment due to unobserved time-variant individual characteristics, we provide evidence on pre-treatment parallel trends and estimate placebo regressions. Although selection dependent on unobserved time-variant individual characteristics may remain, the institutional setting supports our identification. Moreover, we complement our main analysis with several robustness checks. Overall, our main results are robust to a variety of specifications and samples.

We obtain four main results. First, in line with the theoretical predictions, we find that the deterrence effect of tax audits is positive and statistically and economically

² However, in this latter study, 60% of audited taxpayers have no income from self-employment, which is the type of income for which no third-party information is available and therefore tax audits have a more relevant role.

³ DeBacker et al. (2015) study the impact of operational audits on corporations. However, the tax behavior of corporations is likely to depend not only on deterrence and intrinsic motivations, but also on incentives provided within the agency contract between the principal (i.e., the shareholders) and the agent (i.e., the managers) (Slemrod, 2004).

significant. Reported income increases on average by approximately 8.4% after audit. Second, we find a lasting impact of tax audits. Third, we find no effect of audits on subsequent tax reporting in case of null-outcome audits. Finally, our results show that high reported-income taxpayers exhibit a lower change in tax compliance after an audit compared to low reported-income taxpayers, in line with the finding of Slemrod et al. (2001). Altogether, our results tend to reinforce the view that after an audit taxpayers update upward the perceived probability of being audited and the perceived probability that undeclared income is detected conditional on audit, and respond to such update rationally by increasing compliance.

The paper proceeds as follows. Section 2 describes the Italian institutional background and the tax reporting and auditing scheme. Section 3 presents the data. Section 4 describes the empirical methods we employ. Section 5 contains estimates of the impact of audits on subsequent tax behavior. Section 6 provides robustness checks. Section 7 presents a back-of-the envelope calculation of the net benefit of audits. Section 8 draws some conclusions and indicates directions for future research.

2 Tax administration in Italy

In Italy, individual taxpayers are required to pay taxes yearly on all personal incomes earned in each tax year. The tax year aligns with the calendar year. Incomes earned in a given tax year have to be reported between May and September of the following calendar year. Personal incomes may derive from dependent work, self-employment, sole proprietorship and capital (shares in a partnership or in a corporation).

After incomes are declared, tax reports can be audited. The Italian revenue agency (*Agenzia delle Entrate*, henceforth AE) can audit tax reports for up to five years (ordinary expiration period) after the end of the calendar year to which the declaration refers. Then, after five years, evasion can no longer be prosecuted unless it is the outcome of a fraud or a criminal act, in which case the expiration period may be longer.

As regards the audit policy, each year the central directorate of the AE sets specific targets in terms of the number and types of taxpayers to be audited. It uses various public and private databases to assess audit risk. However, both regional and provincial directorates possess a given degree of autonomy in the selection of the taxpayers to be audited. They can conduct their own risk assessment, based on information collected on the spot, and identify risky taxpayers not selected at the central level. Business Sector Studies (*Studi di Settore*, SDS, see Santoro & Fiorio, 2011) are an important source of information for the AE. For each taxpayer subject to SDS, a presumptive revenue (i.e., value of sales) is computed by multiplying the input values (reported by the taxpayer) by the input productivity as computed by the AE. This latter is obtained by regressing revenues on input values reported by a subset of taxpayers belonging to the same cluster as the taxpayer under consideration

and who are classified as reliable by the AE.⁴ In contrast, costs are not directly considered during this process.

Audits generate an audit notice which contains the preliminary tax adjustment claimed by the AE.⁵ The taxpayer becomes aware of such adjustment only when she receives the audit notice. This initial adjustment is based on preliminary desk-type auditing activities that the AE conduct before sending the audit notice, such as cross-checking of accounting and tax data, comparison of reported data with presumptive values and with data from bank accounts.

The preliminary adjustment, however, does not complete the audit process, since the taxpayer can react in different ways. First, in case of a formal mistake, the taxpayer can complete a form asking the AE to repeal the audit notice providing supplementary information to motivate such a claim. Based on new information, the AE can repeal the legal act and formally declare that the taxpayer is compliant (with respect to the specific claim that occasioned the assessment). We call this case ‘null-outcome audit’ (*annullamento per autotutela*). Second, the taxpayer can accept the tax adjustment settled by the AE and pay the due amount with no sanctions within a given deadline (*acquiescenza*).⁶ If she does not, the audit process may continue in two ways. The first is a demand for a settlement, whereby the AE and the taxpayer engage in a sort of bargaining process and the AE gives up a part of the positive adjustment.⁷ The taxpayer immediately pays the tax debt and a reduced sanction, so that the AE saves on administrative costs related to tax collection (*accertamento con adesione*). The second way in which the process may continue is through a legal dispute against the audit note, whereby the case is brought before a special tax court (*contenzioso*). The procedure can last for various years before being legally established. The final outcome, which we do not observe in the data, can be a total or partial cancelation of the preliminary assessment. Figure 1 summarizes the full range of audit outcomes.

2.1 Timing of audit’s effect

According to the AE definition a ‘year t ’ audit is an audit carried out between July 1st of year $t - 1$ and June 30th of year t . Taxpayers are immediately notified when an audit is opened.⁸ A ‘year t ’ audit overlaps with two tax years ($t - 1$ and t) and with

⁴ The definition of “reliability” is not disclosed by AE, so that no collusion can affect the input productivity estimates.

⁵ Note that an audit notice can refer to multiple taxes (for example, to both income and value added taxes), but in this paper we consider only adjustments referring to personal income tax.

⁶ Such deadline is set up to 60 days after the receipt of the audit notice.

⁷ The taxpayer has to set up the procedure within 15 days from the audit notice.

⁸ This is undoubtedly a great advantage of our data with respect to other papers. DeBacker et al. (2015), for instance, observe the moment when an audit is opened and closed, but they do not know when the taxpayer is notified of the audit. They are therefore unable to observe precisely when taxpayers become aware that the treatment has been assigned and, accordingly, update the subjective detection probability. In our case, instead, we are sure that taxpayers can react to the treatment starting from the first tax return filed after an audit. On the other hand, we do not observe the exact date on which the audit is conducted, but only the interval, i.e., the audit year as previously defined.

two tax reports (referring to tax years $t-2$ and $t-1$, respectively). Note that tax reports referring to year t are issued between May and September of year $t+1$, thus after a 'year t ' audit. Figure 2 summarizes the reporting and auditing time structure.

It is highly unlikely that a 'year t ' audit has an impact on reports for tax year $t-2$, since this would require an audit to be conducted between July and September of year $t-1$, which rarely occurs. In contrast, a 'year t ' audit may have an impact on reports for tax year $t-1$, since the corresponding report has probably been issued after the audit. However, tax year $t-1$ may be already concluded, and this limits the adjustment margins for the taxpayer. Note that the taxpayers we observe usually adopt a simplified accounting system that is of limited relevance for tax purposes. This means that, even after the tax year has concluded, tax reports can be modified and not fully adhere to the accounting registrations actually made during the year. Thus, some incomes not recorded in the accounting books can emerge later in the tax report. Finally, a 'year t ' audit is very likely to have an impact on tax reports referring to tax year t , since the audit is conducted before the tax year has concluded and thus before the tax report referring to tax year t is issued.

In view of the audit's above-described temporal structure, we will check for the impact of a 'year t ' audit starting from reports referring to tax year $t-1$, although we expect a lower impact in the first year since the audit may be conducted when the tax year has already concluded.

3 Data

We analyze a panel of Italian taxpayers using data from two sources, both released by the AE. The first dataset contains information from the Tax Return Register "*Anagrafe Tributaria*", which includes the tax reports of all Italian taxpayers. The available sample comprises the universe of VAT registered taxpayers with legal residence in three of the most populated Italian regions, namely Lombardy (located in the North), Lazio (located in the Center) and Sicily (located in the South), which account for around one third of the entire Italian population. VAT registered taxpayers usually obtain their income mainly from two judicial forms: self-employment (*lavoro autonomo*, e.g., single professionals like lawyers or architects) and sole proprietorships (*impresa individuale*, e.g., retailers or handicraft workers).⁹ The sample includes 528,540 taxpayers observed for the 2007–2011 period, corresponding to 2,642,700 observations and to one third of the Italian population of taxpayers with income obtained mainly from self-employment and sole proprietorships. Due to very extreme income values in the lower and upper tails of the income distribution,

⁹ A VAT registered taxpayer, in a given year, may report (also) other types of income. For example, incomes that professionals obtain as members of company boards are treated, for tax purposes, as incomes from dependent work. This motivates our choice to take, as our main dependent variable, reported income obtained from self-employment and from sole proprietorships.

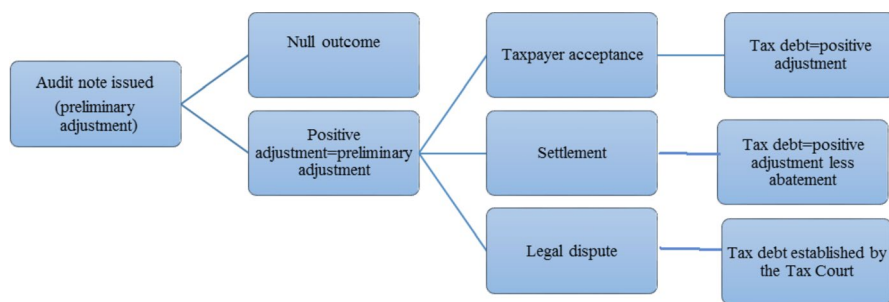


Fig. 1 Audit's outcomes

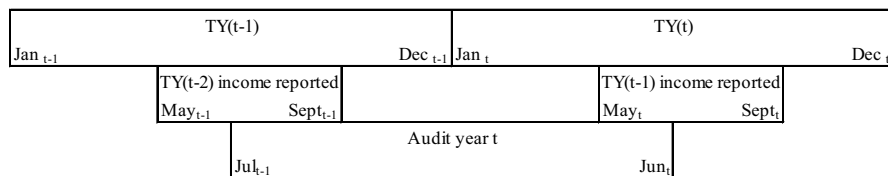


Fig. 2 Reporting and auditing time structure. *Notes:* ty: tax year

we drop observations belonging to extreme positive and negative percentiles to deal with such outliers.¹⁰

The kind of taxpayers included in our sample have a high opportunity to evade (Cabral et al., 2014; Pissarides & Weber, 1989; Slemrod et al., 2001) because most of their incomes are not subject to third-party reporting (Kleven et al., 2011).¹¹

The tax return dataset contains information on a set of taxpayers' demographic characteristics, like gender, age and place of residence, as well as on the main characteristics of taxpayers' economic activity, like the sector and the number of dependent workers. It includes a range of tax-related variables taken from tax returns, like income type (from self-employment or sole proprietorship), incomes from various sources, personal income tax base, gross tax, total amount of tax allowances, net tax. The tax return dataset also contains information related to the implementation of SDS, in particular the presumptive revenues and whether taxpayers' revenues are higher than the presumptive value.¹² Specifically, the dataset provides two indicators based on SDS, namely whether a taxpayer is coherent,

¹⁰ After this censoring, reported incomes range between around -223k and +220k euros. We perform our analysis dropping outliers because we prefer to adopt a more conservative approach. Indeed, estimates of the audits' effect obtained on the restricted sample are lower but presumably also more reliable compared to estimates on the full sample (available upon request).

¹¹ Although recent research has argued against the belief that employees do not evade (Best, 2014; Paulus, 2015), Kleven (2014), using data for more than 80 countries, shows a strong negative relationship between the tax take and the share of self-employed workers, consistent with the view that the coverage of third-party reporting is a crucial determinant of tax enforcement.

¹² Unfortunately, the tax audit dataset does not contain reliable information on reported revenues.

i.e., her input values are internally consistent, and congruous, i.e., her turnover is consistent with input values. However, this information is not available for around 20% of taxpayers because the SDS do not apply. Taxpayers with missing values on these variables are not randomly selected because taxpayers may self-qualify as SDS non-applicants by claiming to be in a 'non-normal' situation or may manipulate the value of presumptive revenue (Santoro & Fiorio, 2011), and the available data do not allow us to distinguish these cases.

The second source of data is the tax audit database. For each audit, it contains information on the amount of the preliminary adjustment, the audit year and the outcome of the audit, distinguishing among null outcome, no taxpayer reaction, settlement, and legal dispute.

The tax return and the tax audit dataset are merged using an encoded taxpayer number (to ensure anonymity) and the tax year. In our sample period, 21,095 taxpayers have been audited at least once (audit rate 4%). Over 96% of audited taxpayers (20,307 taxpayers) have received only one audit and, among them, the vast majority (17,351 taxpayers) have been audited on one tax return. Only 774 taxpayers have received two audits in the sample period and also in this case for the vast majority of them (638) each audit has been conducted on one tax return. Among the very few taxpayers audited three times (13), each audit has been conducted on one tax return with the exception of two taxpayers. Finally, only one taxpayer has been audited four times and on four tax returns (i.e., one for each audit).

In the following empirical analysis, we will focus on taxpayers audited once and on a single tax return, that we label 'single-audit' taxpayers. We believe that the process of selection of taxpayers audited more than once and/or on more than one tax report is more likely to be driven by non-observable and time-variant individual characteristics that we are not able to control for and that would bias our estimates of the audit effect. In addition, taxpayers that have been audited more than once in the sample period or for whom a single audit has inspected more than one tax return received a different treatment, namely a more intensive one compared to taxpayers audited once and on a single tax return. Moreover, our analysis of audit's effect by audit's outcome would be problematic for taxpayers receiving more than one audit with different outcomes. Finally, with regard to taxpayers audited more than once, on average the after-audit period is longer than in the single audit case (2.5 vs 1.5 years). This would affect our results because, as we will show, the dynamic audit effect is not linear. Overall, we believe that by focusing on single-audit taxpayers we select a more homogeneous sample for which concerns regarding selection into audit are lessened.

Table 1 shows that almost 90% of the audits that we observe are conducted in 2011 (28%) or 2012 (62%). The reason is related to the nature of our data and to the Italian institutional setting regarding tax audits. By law, tax returns older than five years cannot be audited, so a full auditing cycle lasts 5 years. Then, tax returns reporting 2007 (t) income can be audited from 2008 ($t+1$) to 2012 ($t+5$). Moreover, since a tax audit conducted in year t affects tax returns of year $t-1$, only audits

conducted up to 2012 may affect tax returns in our dataset. We then focus on audits conducted between 2008 and 2012.¹³

Since tax returns older than five years cannot be audited anymore, the Revenue Agency has an incentive to audit tax returns close to the expiration date in order not to lose the possibility of auditing them. The consequence of this is that most audits on 2007–2011 tax returns (i.e., those observed) will be conducted in 2011 and 2012 on 2007 and 2008 tax returns.¹⁴ For instance, considering audits carried out in 2012 (for which we can observe the entire period over which audit can be conducted, i.e., 2007–2011), 60.7% refer to the 2007 tax return and 30.1% to the 2008 tax return.

As regards the audit outcome, 10.8% of single audits end up with a null outcome, while the remaining 89.2% have a positive adjustment.

One weakness of our data is that they cover only 5 years. This implies that taxpayers that have not been audited in our sample period could have been audited just before (e.g., in 2006). Hence, their behavior in the 2007–2011 period (especially in the first years of the period) may be influenced by audits that we are not able to observe. We will perform a robustness check in order to address this limitation.

A second limitation is that our data contain only observations relative to taxpayers present continuously in the tax audit dataset for the whole 2007–2011 period, while we do not observe taxpayers who entered or left the register because they either started or closed their business (e.g., due to death or to closure caused by bankruptcy). On average these latter may react to an audit differently from the taxpayers in our sample. For instance, taxpayers close to bankruptcy may fail to react because of a cash flow shortage. Notice, however, that a time-variant sample composition would bias our analysis of the dynamic audits' effect.¹⁵

To analyze the effects of enforcement actions on subsequent reporting behavior, we consider as outcome the total personal income from self-employment and sole proprietorship as reported on taxpayers' tax returns.¹⁶ Taxes on this type of income are highly subject to evasion through misreporting or underreporting income. However, given that the magnitude of and the opportunity for evasion differ widely across types of income and deductions, we will use other outcomes in the robustness check Section. Specifically, we will check if our results hold when considering total before-tax income, taxable income, and net tax.

¹³ Actually, we observe audits conducted up to 2013 but 2013 audits do not affect the filing of the tax returns in our dataset (2007–2011).

¹⁴ On the other hand, 2010/2009/2008 audits will regard mainly, respectively, 2005–2006/2004–2005/2003–2004 tax returns that we do not observe in our data.

¹⁵ In view of this concern, DeBacker et al. (2015), despite the availability of the entire population of tax filers, drop companies entering and exiting during the sample period to avoid this type of survivorship bias.

¹⁶ The Italian personal tax on income is computed by applying the increasing marginal rates of the progressive tax schedule to the taxable income, which is obtained by summing the different sources of income.

Table 1 Distribution of audits by audit year and by inspected tax return

Audit year	Tax return					
	2007	2008	2009	2010	2011	Total
2008	64	–	–	–	–	64
2009	488	15	–	–	–	503
2010	1112	179	9	–	–	1300
2011	2564	2088	119	14	–	4785
2012	6495	3224	756	208	16	10,699
Total	10,723	5506	884	222	16	17,351

Figures in the table refer to the matched sample of taxpayers audited once and on one tax return

4 Methods

In the first part of this Section, we discuss the identification issues related to the estimation of the causal effect of real-world operational audits on subsequent tax compliance, and we present graphical evidence supporting our identification strategy. In the second part, we illustrate the estimated equations, and in the final part we present a test of the common trends assumption along with some placebo regressions.

4.1 Identification issues

When estimating the extent to which taxpayers adjust tax compliance behavior in response to an audit, we have to consider that audits (the treatment) are unlikely to be randomly assigned. In general, revenue agencies tend to audit subjects with a higher expected net return on the audit, maximizing the difference between its expected benefits and costs. This selection process can lead to biased estimates of the causal effect of audit on tax compliance, because of correlation between selection and tax compliance.

A first threat to identification is that the choice of the subjects to be audited may be based on time-invariant taxpayers' characteristics that are likely to be correlated with the outcome (reported income) but that we may not be able to observe. The panel structure of our data, with information on both pre-treatment and post-treatment periods, enables us to circumvent this obstacle to identification. Our identification strategy relies on fixed-effect difference-in-difference comparison with non-audited taxpayers. We compare changes in outcome between taxpayers who were audited (the treated group) and taxpayers who were not audited (the control group). Moreover, considering that audited and non-audited taxpayers may differ in both observable and unobservable characteristics, we use an ex-ante approach to restrict our control sample and increase the similarity between treated and control groups. Specifically, we use exact matching: for each taxpayer exposed to the treatment (i.e., audited), we identify unaudited taxpayers that match exactly based on gender, industry (classification based on 21 NACE groups), province (i.e., the geographical level

at which the auditing policy is mainly established), age deciles and income quartiles in 2007, which is the beginning of our period of analysis.¹⁷ Given the large control sample available, we are able to match almost all audited taxpayers to one or more non-audited taxpayers, and we drop from our sample the very small number of audited taxpayers (294) that do not have a similar counterpart in the non-audited individuals. As robustness checks, we will show results obtained using a different set of matching variables and using the unrestricted sample as well.

Overall, in our sample we have 16,741 treated taxpayers that were audited once and on one tax return in any of the 2007–2011 years and that have at least one counterpart in the control group. Considering that the matching algorithm can match the same untreated taxpayer to more than one audited taxpayer, after unmatched observations are dropped the control group is composed of 367,156 distinct taxpayers. Summary statistics for audited and non-audited groups (both matched and non-matched) are reported in Table 2 for 2007. The means for the two groups are very close as regards age and gender distribution. Audited taxpayers are relatively less concentrated in Lombardy, the northern region, than in both the Center and Southern regions. Audits are relatively more frequent in specific industries. Considering the most represented industries, audits occur more often in the wholesale and retail, transport, accommodation and food service activities. The average pre-treatment income from professional and firm activity, the gross income, the taxable income and the average net tax paid are slightly higher for audited taxpayers.¹⁸ Column 2 shows descriptive statistics for non-audited matched taxpayers computed using coarsened exact matching (CEM) weights.¹⁹ Comparing matched and non-matched non-audited taxpayers with audited taxpayers, it can be seen that matching marginally improves the balancing of the income variables.

A second threat for identification is that time-variant individual characteristics may affect treatment assignment. In particular, one may be concerned that audits are concentrated in years when reported income is relatively low or high. For instance, if the AE considers low reporting as a sign of evasion, an audit may be carried out when, for a given taxpayer, reported income is low relative to her average reported income. This could invalidate our empirical design, and we would obtain upward biased estimates of the audit effect due to mean reversion. A similar concern is that taxpayers to be audited are selected by the AE looking at the value of other

¹⁷ Estimation that combines ex-ante matching on observable characteristics with fixed-effects to account for time invariant unobserved factors produces more reliable estimates than matching alone (Smith & Todd, 2005).

¹⁸ Average income variables are substantially higher when we avoid matching, suggesting that audits are actually not random. Specifically, it seems that the AE tends to audit more often taxpayers with higher levels of average reported income.

¹⁹ CEM weights for control units are defined as the ratio between the number of controls and the number of treated in the sample by the ratio between the number of treated and the number of control in the specific stratum defined by the matching variables. Weights for treated units are equal to 1 (Iacus et al., 2012).

Table 2 Baseline taxpayers characteristics

	Non audited		Non audited (matched)		Audited	
	Mean	Std dev	Mean	Std dev	Mean	Std dev
Outcomes						
Personal income from professional and firm activity	31,023	32,483	31,571	33,925	31,625	37,093
Gross income	30,759	32,638	31,458	33,993	31,565	37,120
Taxable income	31,705	56,671	32,461	56,090	35,480	47,912
Net tax	8339	23,414	8652	23,127	9839	19,249
Age	45.9	11.2	47.1	11.6	47.2	11.5
Male	0.77		0.80		0.80	
Industry share						
Agriculture, forestry and fishing	7.3		3.4		3.4	
Mining and quarrying	0.0		0.0		0.0	
Manufacturing	6.3		6.7		6.7	
Water supply; sewerage; waste manag. and remediation activities	0.1		0.1		0.1	
Construction	12.8		12.3		12.3	
Wholesale and retail; repair of motor vehicles and motorcycles	25.4		30.3		30.3	
Transporting and storage	2.7		3.9		3.9	
Accommodation and food service activities	4.1		7.9		7.9	
Information and communication	1.9		1.1		1.1	
Financial and insurance activities	1.3		1.6		1.6	
Real estate activities	1.1		1.1		1.1	
Professional, scientific and technical activities	24.3		20.1		20.1	
Administrative and support service activities	2.6		3.0		3.0	
Education	0.3		0.3		0.3	
Human health and social work activities	4.3		3.5		3.5	
Arts, entertainment and recreation	1.6		1.9		1.9	

Table 2 (continued)

	Non audited		Non audited (matched)		Audited	
	Mean	Std dev	Mean	Std dev	Mean	Std dev
Other services activities	4.0		3.1		3.1	
Region						
Lombardy (North)	54.1		32.3		32.3	
Lazio (Centre)	25.2		37.4		37.4	
Sicily (South)	20.8		30.4		30.4	
Nr taxpayers	501,041		367,156		16,741	

The matched sample of non-audited taxpayers is restricted to non-audited taxpayers matched to the sample of taxpayers audited once and on one tax return based on gender, industry, province, age and baseline income bracket. We used CEM (coarsened exact matching) weights for the matched sample of non-audited taxpayers. The sample of audited taxpayers is restricted to taxpayers audited once and on one tax return in any of the 2007–2011 years. We have dropped 293 audited taxpayers without a similar (matched) counterpart in the non-audited individuals. Gross income includes other sources of income like those subject to house tax, rental tax or land value tax in addition to income from professional and firm activity. Taxable income is obtained by subtracting tax deductions like compulsory social security contributions from before-tax income

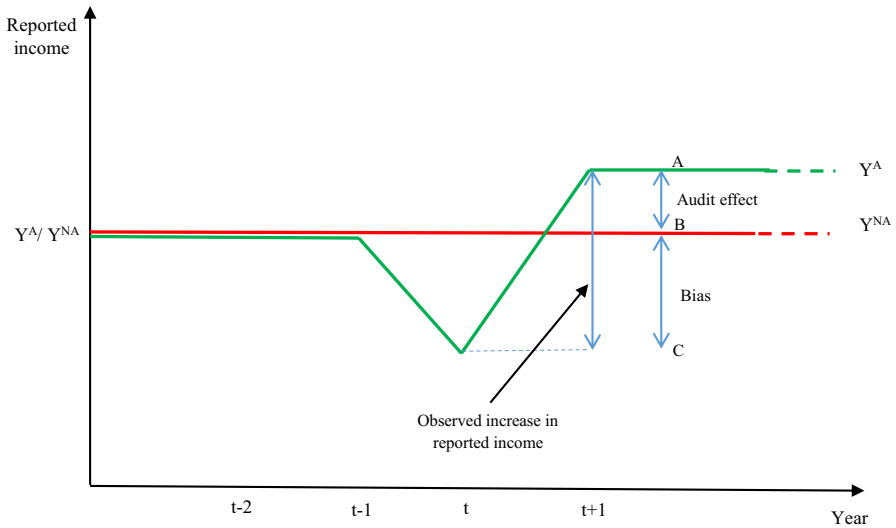


Fig. 3 Estimation bias in presence of time-variant factors affecting treatment assignment. *Notes:* Y^{NA} and Y^A refer to income reported, respectively, by non-audited and audited taxpayer

compliance-related variables such as gross income, taxable income or net tax, which may be time-variant.

The potential bias arising in this case is illustrated through an example in Fig. 3. Let us consider a treated (audited, A) and a control (non-audited, NA) taxpayer. Assume that in period $t-2$ and $t-1$ both taxpayers report a constant and identical income. A temporary drop in taxpayer A reported income (unrelated to changes in her tax compliance) occurs at time t (in the absence of audit, at time $t+1$ reported income would go back to its previous value), while income reported by taxpayer NA does not change. The tax authority considers the drop in reported income of taxpayer A as a signal of evasion and, accordingly, it audits her. Taxpayer A responds to the audit increasing her compliance by the amount AB. However, we would ascribe to the audit all the change in reported income from period t to period $t+1$ (AC in Fig. 3), even if part of this reported income increase (BC) would have occurred also in the absence of the audit.

To address this issue, first of all we consider that both the institutional setting, namely the five-year deadline for carrying out audits (see Section I), and the evidence in our data (see Section II, Table 1) suggest that the main criterion that the AE follows to select the tax return of a given taxpayer to be audited is its closeness to the expiration date. This implies that on average there is a substantial lag between the year of the audited tax return and the year of the treated tax return. Actually, in our data for over 37% of audits such lag is equal to four years (i.e., the year of the audited tax return is 2007 and the audit is carried out in 2012) and for over one third the lag is equal to 3 years (i.e., the year of the audited tax return is 2007 or 2008 and the audit is carried out, respectively, in 2011 or 2012). In view of this, we believe that reverse causality should not invalidate our empirical design because in case it

is a temporary drop of income that triggers audit, it is likely that reported income will go back to its previous average value in a shorter time period compared to the observed lag between audit year and the year of the audited tax return.

To mitigate concerns related to endogenous selection into audits, we focus on taxpayers audited in 2012, for which we observe a complete auditing cycle, and provide additional evidence corroborating our empirical strategy. Specifically, the estimation strategy that we adopt relies on the common trends assumption for untreated periods. We presume that, in the absence of audits, the treated and the untreated taxpayers would have shown a similar trend of tax compliance behavior. Figure 4 provides descriptive evidence for the parallel trends assumption by showing the average reported income and confidence intervals over the period 2007–2011 for audited and non-audited taxpayers. To facilitate comparison of trends, we normalize the lines for audited and non-audited taxpayers so that they take on the same value in 2007.

The figure shows that before the audit average reported incomes for the two groups have similar trends (only in 2008 a very small difference is observed), with average income decreasing for both groups, particularly in 2009 and 2010 when the Great Recession showed its effects. It is also evident that, as we expected, the trends diverge in the post-treatment period (i.e., in 2011), with a much larger increase in reported income for audited taxpayers.

At the end of the following subsection we will provide additional evidence on the validity of the common trends assumption.

A final concern is that in our staggered difference-in-differences set-up, for individuals audited in the first observed years (2007 and 2008), there is no change in the treatment status. In the fixed effects model, they are, hence, included as controls—not as treated units. However, their number is very low (no taxpayer is audited in 2007 and only 64 taxpayers have been audited in 2008). In the robustness section we present results including only taxpayers audited after 2008, for whom there is a change in the treatment status (see Table 7, columns 5).

4.2 Estimating equations

To assess the average impact of audits on subsequent tax compliance, we estimate the relative change in reported income before and after the audit comparing audited taxpayers and the matched control sample of non-audited taxpayers by estimating the following equation:

$$Y_{ijt} = \beta_1 TREATED_i \times Post_{it} + \alpha_i + \tau_t + \sigma_j + \mu_k + \gamma_1 X_{ijt} + \varepsilon_{ijt} \quad (1)$$

where Y_{ijt} measures personal income as reported by taxpayer i in year t in industry j in province k and $TREATED_i$ is a dummy equal to one for audited taxpayers. The variable $Post_{it}$ is a dummy equal to one for each period after taxpayer i becomes treated (i.e., from one year before the audit onwards, see Sect. 3). The effect of audits on tax compliance is captured by β_1 , the coefficient for the interaction term between treated taxpayers and the post-audit period.

The terms α_i , τ_t , σ_j and μ_k are individual, year, industry and province fixed effects, respectively. The individual fixed effects control for any observed or unobserved

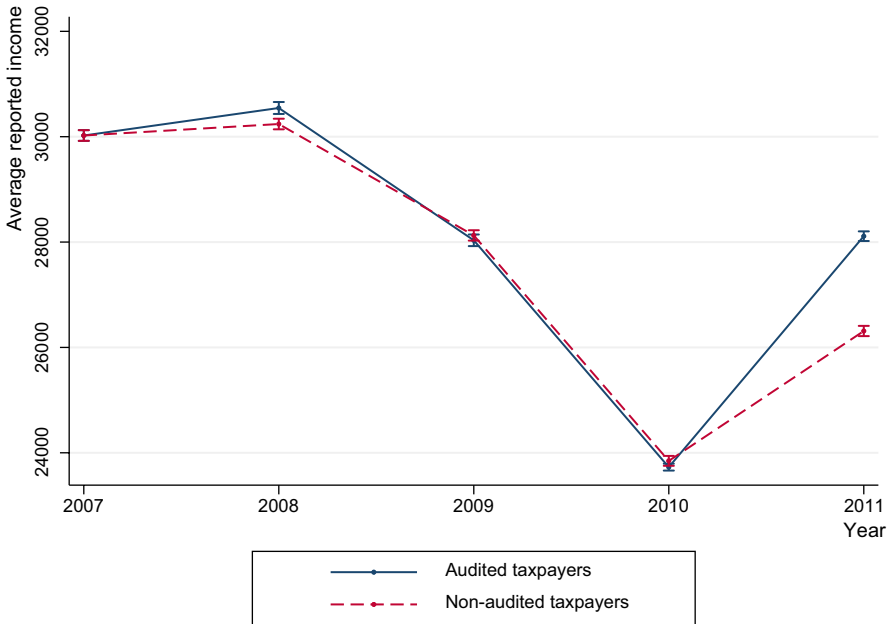


Fig. 4 The common trends assumption. Mean and confidence intervals. *Notes:* The sample of audited taxpayers is restricted to taxpayers audited in 2012 (year after treatment = 2011)

individual characteristics that are constant over time and that may affect the outcome. The year fixed effects control, in addition to macroeconomic fluctuations in general economic activity, for yearly changes in auditing guidelines established at central level. The inclusion of individual, year, industry and province fixed effects should ensure that our comparison across treatment groups over time is not influenced by group-specific characteristics. X is a vector of taxpayer characteristics, including gender, age and its square (that are absorbed by the individual fixed effects when included in the regression), a dummy for “coherent” and “congruent” taxpayer and, in order to deal with missing values, a dummy for taxpayers for which the presumptive revenue is not defined. Finally, ε is an error term.

Our model is very similar to the one used by De Backer et al. (2018), where estimation of the audits’ effect comes from within-individual changes in reported income between post-audit and pre-audit periods, net of trends in income common across the treatment and control groups, which are accounted for by the year fixed effects. This approach is preferable to the standard difference-in-difference approach without individual fixed-effects adopted by D’Agosto et al. (2018), who use a dataset very similar to ours, considering that there are reasons to believe that some time-invariant individual characteristics may influence treatment assignment.

The next step is estimating the dynamic effect of audits. More specifically, we want to assess whether and how the audit’s impact on reported income changes

over time. For this purpose, we extend the primary specification given by Eq. (1) estimating the following equation:

$$Y_{ij} = \sum_{k=0}^4 \beta_k TREATED_i \times D_k + \alpha_i + \tau_t + \sigma_j + \gamma_2 X_{ij} + \varepsilon_{ij} \quad (2)$$

where D_k are a series of dummy variables, one for each year after taxpayer i becomes treated (that is from one year before the audit onwards), and the other variables retain the same meaning as in Eq. (1).

4.3 Additional evidence in support of the identification strategy

In this section, we test more rigorously the parallel trends assumption in the pre-treatment period focusing on taxpayers audited in 2012 for which we observe a full auditing cycle. We use their pre-treatment data from 2007 to 2010 and estimate the following equation (Muralidharan & Prakash, 2013):

$$Y_{ij} = \delta_1 TREATED_i \times Trend + \delta_2 \times Trend + \alpha_i + \sigma_j + \gamma_1 X_{ij} + \varepsilon_{ij} \quad (3)$$

where $Trend$ is a linear variable taking the value of 1 in 2007 and ending in 2010, while other variables are defined as in Eq. (1). Estimation results are presented in Table 3 (column 1); the lack of statistical significance of the coefficient of the interaction term δ_1 suggests the existence of pre-treatment parallel trends. If we use a nonparametric specification using dummies for pre-treatment years rather than assuming a linear trend, we observe a slightly larger decrease in income for audited taxpayers in pre-audit years with respect to the baseline year (2007), although the estimates are small and not statistically significant (see Table 3, column 2).

To corroborate our identification strategy, we estimate two *placebo* versions of Eq. (1) using data from the before-audit period for taxpayers audited in either 2011 or 2012. In the first regression, we compare relative changes in reported income from 2007 (fake pre-treatment) to 2008–2009 (fake post-treatment) for audited taxpayers with respect to the same change in reported income for untreated individuals. We replicate the same exercise including 2008 in the pre-treatment period (then considering 2009 as fake post-treatment period). The results in Table 3 show that the coefficients of the *fake* post-audit variable are small and never statistically significant (columns 3 and 4).

Overall, the previous analysis indicates that there are not substantial differences in income development between audited and non-audited taxpayers in the pre-audit period. In view of this, while we cannot say what would happen to audited taxpayers' income in the absence of audit, we do not believe that mean reversion, implying a substantial change in income development with respect to what we observe in the pre-audit period, is a relevant mechanism for interpreting what we observe in the after-audit period.

5 Results

Below, we provide a series of estimates of the impact of audits on reported income based on Eq. (1). We also test for different audit effects along the reported income distribution and we estimate the dynamic impact of audits based on Eq. (2). Afterwards, we present audit effects by audit outcome. In all specifications, standard errors are clustered at taxpayer level. We deflate all nominal values to 2011 euro using the CPI to have income variables in real terms.²⁰

5.1 Average audit effect

Estimates of average audit effects based on Eq. (1) are reported in Table 4. In column 1, we show OLS results including only the “*Treated* × *Post*” variable and year fixed effects; the specification shown in column 2 includes also industry and province fixed effects and controls for individual-level variables; columns 3 and 4 show results from different specifications of the difference-in-difference fixed-effect model.

The coefficient β_1 of the “*Treated* × *Post*” variable is positive and highly statistically significant in all columns. On comparing point estimates of OLS regressions, we find a higher coefficient when more controls are entered in the regression. More specifically, the increase in β_1 occurs when the provincial and the sectoral dummies are added. This is explained by the relatively greater concentration of audits in southern provinces of Sicily, where tax evasion is higher and the average reported income is considerably lower than in other regions. Data also indicate that on average reported income is lower in more frequently audited industries.

When we use the fixed effect estimator (columns 3 and 4), then controlling for time-invariant individual characteristics, the positive impact of audits is confirmed. We find that annual reported income grows on average by around 2.5 thousand euro after receiving an audit, corresponding to around 8.4% of audited taxpayers’ average reported income.²¹

Overall, the positive impact of audits on reported income is consistent with the prediction of the expected-utility maximization model of tax compliance that audited taxpayers increase future compliance when they revise upward their perceived probability of being audited.

To scrutinize further audit effect, we investigate whether it varies along the reported income distribution. To do so, we estimate Eq. (1) interacting the post-audit

²⁰ A similar approach is followed for instance by DeBacker et al. (2015). Our results are unchanged if we forgo deflating and make the analysis in nominal terms.

²¹ Notice that fixed effect estimate of the audit’s effect is lower than the OLS value. Although we use income quartiles as matching variable, this may be related to the selection by income to the audit group within quartiles and to the difference in development of higher versus lower individuals. Indeed we find that if we use the logarithm of reported income as dependent variable, OLS and fixed effect estimates are much more similar. In the robustness check, we use the logarithm of reported income as dependent variable.

Table 3 Test of the parallel trends assumption and placebo regressions

Variables	(1) Test parallel trends (linear trend)	(2) Test parallel trends (year interactions)	(3) Placebo (2007/2008– 2009)	(4) Placebo (2007– 2008/2009)
Treated × Trend	– 156.6 (114.058)			
Treated × 2008		– 211.892 (239.429)		
Treated × 2009		– 468.500 (294.163)		
Treated × 2010		– 434.860 (350.563)		
Audit effect (placebo)			20.467 (189.238)	106.515 (198.634)
Individual controls	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Observations	1,506,164	1,506,164	1,143,610	1,143,610
Number of id	377,481	377,481	382,094	382,094
R-squared	0.064	0.069	0.039	0.039

Robust standard errors are in parentheses, clustered at the taxpayer level. The audited taxpayer sample is restricted to taxpayers audited in 2012 (columns 1 and 2) and in either 2011 or 2012 (columns 3 and 4). Sample period is 2007 to 2010 (columns 1 and 2) and 2007 to 2009 (columns 3 and 4). Column 1 reports estimates of Eq. (3). Column 2 reports estimates of Eq. (3) using a nonparametric specification with dummies for pre-treatment years rather than a linear trend. Columns 3 and 4 report estimates of Eq. (1) with 2 different fake after audit periods. Individual controls are a dummy for “coherent” or “congruent” taxpayer and a dummy for “non-coherent” and “non-congruent” taxpayer (the omitted category includes taxpayers for which the presumptive revenue is not defined and for which, accordingly, the “coherent” and “congruent” variables have missing values)

*Significant at 10% level

dummy with income deciles computed at the beginning of the period.²² Moreover, to take into account that income deciles in 2007 may be not a good measure for income of subsequent years, and to be sure that audited and non-audited individuals are still comparable considering that audits are more frequent among large income taxpayers, we use decile defined in 2008 instead of 2007 and we use quartiles defined in both 2007 and 2008 instead of deciles as well.²³

²² Descriptive statistics for income deciles are reported in Table 10, Appendix 2.

²³ We use 2007 and 2008 and not later years because in these years almost all audited taxpayers are pre-treatment and we prefer to create deciles using incomes not including audit’s effect.

Table 4 Average audit effect

	Reported income			
	(1)	(2)	(3)	(4)
	OLS w/o control	OLS with controls	FE w/o control	FE with control
Audit effect on subsequent reporting	1489.871*** (271.104)	3941.869*** (243.534)	2687.717*** (190.996)	2543.675*** (188.172)
Constant	32,533.072*** (55.065)	- 31,322.816*** (524.978)	32,300.752*** (26.730)	- 7903.337*** (1270.059)
Individual controls	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes
Province FE	No	Yes	No	Yes
Industry FE	No	Yes	No	Yes
Individual FE	No	NO	Yes	Yes
Observations	1,915,186	1,915,186	1,915,186	1,915,186
Number of id			384,213	384,213
R-squared	0.009	0.244	0.035	0.061

Columns 1 and 2 report OLS estimates. Columns 3 and 4 report fixed-effect estimates. Robust standard errors are in parentheses, clustered at the taxpayer level. Individual controls are gender, age, squared age (that are absorbed in the fixed-effects estimates in columns 3 and 4), a dummy for “coherent” or “congruent” taxpayer and a dummy for “non-coherent” and “non-congruent” taxpayer (the omitted category includes taxpayers for which the presumptive revenue is not defined and for which, accordingly, the “coherent” and “congruent” variables have missing values)

***Significant at 1% level

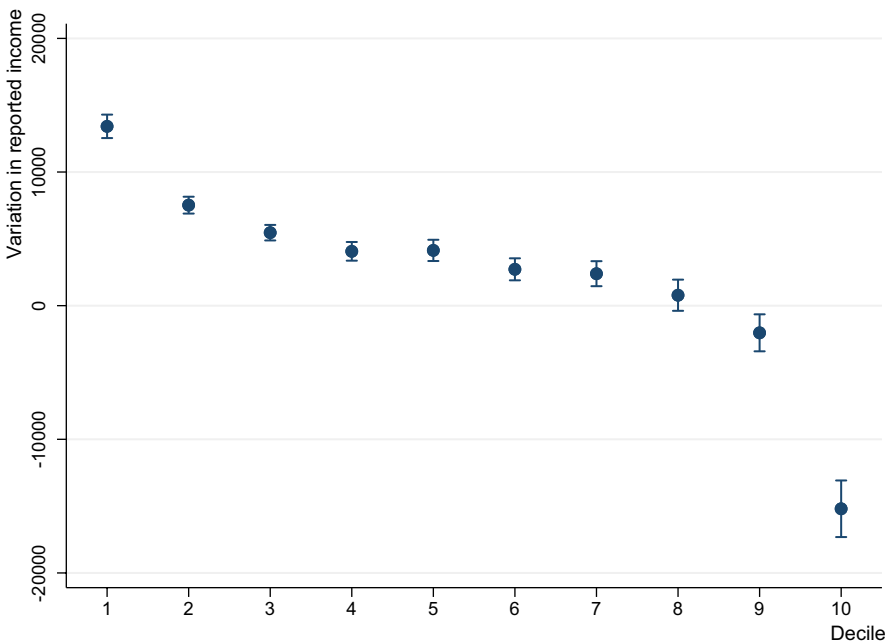


Fig. 5 Audit effect by reported income decile

The results with deciles computed at the beginning of the period are reported in Fig. 5.²⁴ They clearly show that the audit effect is higher at the lowest deciles. Incidentally, the first decile includes only negative and zero income taxpayers and it appears that they react considerably more compared to other taxpayers. One potential explanation is that such taxpayers represent high wealth filers who are very tax aggressive. Unfortunately, however, with our data we cannot test this interpretation. In general, the β_1 coefficient decreases monotonically along the reported income distribution, turning negative (minus 15.2 thousand euro) at the last decile. Our results of audit's effect decreasing in income is confirmed when using deciles defined in 2008 and quartiles instead of deciles (see Table 9, Appendix 1). This result suggests that the average positive audit effect that we detected before is driven by low and middle reported-income taxpayers, while the effect is even negative at the highest decile. A similar result is found by Slemrod et al. (2001) in the Minnesota experiment, where a group of randomly selected taxpayers were informed by letter that the tax returns that they were about to file would be audited. They found that high-income taxpayers report less when they expect an audit.²⁵ The main explanation provided by Slemrod et al. (2001) is that high-income taxpayers tend to believe that the final outcome of an audit depends on the initially reported income and that an audit will not necessarily discover all evasion. This belief is based on the assumption that high-income individuals are more likely to receive professional assistance with their tax affairs. We cannot test this hypothesis because reliable information about the presence and type of tax consultant is not available in our data.

Another potential interpretation is that higher income individuals may have higher tax compliance and they react less because they have a lower tax gap.²⁶ Alternatively, this result may be related to differences in the marginal cost of increasing compliance along the income distribution due to the personal income tax progressivity. For instance, the marginal cost of reporting more income after the audit (the marginal effective tax rate) may be small or even zero for taxpayers whose taxable income is negative even after the post-audit correction.²⁷ Unfortunately, however, with our data we are not able to test rigorously the validity of the proposed interpretations.

5.2 Dynamic audit effect

It is possible that the after-audit tax behavior changes over time because, as time since audit goes by, taxpayers may revise their subjective audit probability based on

²⁴ Full estimation results and results using the other specifications are in Table 9, Appendix 1. In these estimates we do not separate the bottom decile among the negative/zero and positive income groups.

²⁵ Actually, our results are not directly comparable to those in Slemrod et al. (2001) because in this latter study the treatment is the audit letter, while in our case it is the audit itself.

²⁶ In our data the size of adjustment for taxpayers in the highest decile is 40% of the income in the audited tax year, while, for taxpayers in the ninth deciles, the size of the adjustment is more than 85% of the audited reported income.

²⁷ However, on the basis of the results in Kleven et al. (2011) showing a small effect of the marginal tax rate on tax evasion, we do not believe that our result can be explained mostly by this mechanism.

more recent audit experience. More specifically, according to the target effect, individually perceived audit probability should decline with time and, in turn, tax compliance should progressively decrease. Through the dynamic analysis in this section, we test for the existence of different after-audit effects depending on the time elapsing since the occurrence of the audit. The dynamic effect of audits is represented by the vector β_k in Eq. (2), containing the coefficients of the interactions between the treatment dummy and dummy variables for the number of years elapsed since the audit. Potentially, we observe taxpayers up to 5 years after an audit. However, since the number of taxpayers observed 4 and 5 years after audit is very low because taxpayers are rarely audited just a few months after submitting their tax report, we use a single dummy variable equal to one if the taxpayer has been audited either four or five years before.

In the first year following an audit, taxpayers increase their reported income on average by around 2.1 thousand euro (see Table 5). In the following two years, the audit's effect is higher. The size effect is lower in the two last years although, due to a low number of observations, the coefficient is very imprecisely estimated.

The lower first year effect is probably related to the temporal structure of the auditing mechanism (see Fig. 2): year t audit may occur starting from July of tax year $t-1$, when half of the tax year $t-1$ has already passed. This means that behavior can be changed in response to the audit only in the second part of the fiscal year because, for the first part of the year, tax behavior with tax consequences has already been carried out.²⁸ Moreover, a 'year t ' audit could be carried out between January and July of year t , after tax year $t-1$ has already concluded. Although the Italian law allows for some ex-post upward correction of incomes resulting from accounting books, in this case the possibility to increase compliance is further reduced. For the subsequent tax years, instead, audited taxpayers have more possibility to adjust their behavior. Notice that, given that for 60% of cases we observe only one year after audit, an implication of the smaller first year effect is that our average estimate is a lower bound.

5.3 Audit's effect by audit's outcome

Our dataset is unique in allowing us to investigate audit impact by audit outcome. We distinguish between null-outcome audits and positive-assessment audits separately for high (i.e., above the median level) and low adjustment.

The results are set out in Table 6. They show that in the case of null outcome the coefficient is very small and statistically non-significant. This finding is consistent with the predictions of the standard model of tax compliance, where tax compliance depends on detection probability that, in turn, is the product of the probability of audit and the probability of detection conditional on audit (Kleven et al., 2011). The

²⁸ For many economic activities, accounting procedures and VAT declaration procedures referring to tax year $t-1$ have to be completed within March of the year t (while income tax reports are issued from May until September of year t). Therefore, the earliest audit of audit year t , which occurs in July of year t , follows the completion of accounting procedures and VAT declaration.

Table 5 Dynamic audit effect

	FE
I year effect	2107.9*** (196.8)
II year effect	3972.3*** (317.4)
III year effect	3277.6*** (662.4)
IV and V year effect	1384.1 (1235.9)
Constant	-7808.4*** (1270.6)
Individual controls	Yes
Year FE	Yes
Province FE	Yes
Industry FE	Yes
Individual FE	Yes
Observations	1,915,186
Number of id	384,213
R-squared	0.061

Robust standard errors are in parentheses, clustered at the taxpayer level. Individual controls are a dummy for “coherent” or “congruent” taxpayer and a dummy for “non-coherent” and “non-congruent” taxpayer (the omitted category includes taxpayers for which the presumptive revenue is not defined and for which, accordingly, the “coherent” and “congruent” variables have missing values)

***Significant at 1% level

null effect of audit might result from a combination of two opposite mechanisms: on the one hand, the null-outcome emerges when, following the information provided by the taxpayer, the Revenue Agency decides to repeal the act. This means that the investigative activity conducted by the AE, before issuing the preliminary adjustment, was fundamentally flawed. The knowledge of such a failure by the taxpayer can induce her to revise downward her expectations about tax inspectors’ ability, then, her subjective probability of detection conditional on audit. On the other hand, the fact of having been selected for an audit may induce an upward revision of taxpayers’ prior about the probability of audit. On average, therefore, the effect on future reported income should be lower than in the positive-adjustment case.²⁹ Indeed, our results show that when the outcome of the audit is a positive assessment of additional income, subsequent tax compliance increases significantly for both high- and low adjustment groups. Moreover, taxpayers with larger adjustment react more to audits (also in relative terms) compared to low-adjustment taxpayers.

²⁹ Clearly, this reasoning is based on the expected utility model, i.e., it is not considering the inherently honest taxpayer whose behavior is not influenced by the outcome of the audit.

Table 6 Audit effect by audit outcome

Audit effect	FE
Null audit	- 88.9 (602.8)
Positive assessment audit (high)	3603.0*** (285.9)
Positive assessment audit (low)	1895.7*** (248.6)
Constant	- 30,414.7*** (1210.4)
Individual controls	Yes
Year FE	Yes
Province FE	Yes
Industry FE	Yes
Individual FE	Yes
Observations	1,915,186
Number of id	384,213
R-squared	0.048

Robust standard errors are in parentheses, clustered at the taxpayer level. Individual controls are a dummy for "coherent" or "congruent" taxpayer and a dummy for "non-coherent" and "non-congruent" taxpayer (the omitted category includes taxpayers for which the presumptive revenue is not defined and for which, accordingly, the "coherent" and "congruent" variables have missing values). Positive assessment is high (low) if the assessment is above (below) the median value

***Significant at 1% level

6 Robustness

In this section we explore the robustness of our results. First, we estimate Eq. (1) considering alternative measures of tax compliance as outcomes. Second, we replicate estimates on different subsamples of audited taxpayers. Third, we use the logarithm of reported income as dependent variable. Fourth, we estimate Eq. (1) applying CEM weights. Finally, we present results obtained on the full sample of untreated taxpayers (i.e., without ex-ante matching) and using a different set of matching variables.

As alternative measures of tax compliance, first we consider gross before-tax income. In addition to income from professional and firm activity, gross income includes other sources of income like those subject to house tax, rental tax or land value tax. Next, we test our main results using as outcome taxable income (obtained by subtracting tax deductions like compulsory social security contributions from before-tax income) and the value of the net tax. The estimates in Table 7 (columns 1 to 3) confirm that, for any outcome considered, tax compliance increases in the post-audit period.

Table 7 Alternative outcome variables, samples and estimation methods

	Reported gross income	Taxable income	Net tax	Reported income		Log income	CEM weights	Non matched sample	Matching on income deciles
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
Audit effect	1631.166*** (181.538)	1221.206*** (214.173)	431.837*** (88.848)	1956.366*** (245.464)	2596.124*** (184.367)	0.097*** (0.023)	1816.652*** (191.415)	2399.121*** (186.595)	2458.402*** (187.783)
Constant	-45,966.269*** (1180.506)	-22,372.790*** (2035.467)	-10,732.902*** (854.231)	-7443.742*** (1273.509)	-7762.505*** (1269.093)	14.798*** (0.193)	-5162.517** (2127.903)	-5600,298*** (1076.513)	-12,075.751*** (1435.257)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,915,186	1,915,186	1,915,186	1,882,349	1,914,098	1,869,588	1,915,186	2,597,510	1,505,443
Number of id	384,213	384,213	384,213	377,481	383,851	383,972	384,213	523,051	302,139
R-squared	0.035	0.006	0.003	0.062	0.061	0.112	0.060	0.058	0.060

In columns 1 to 3 the outcome variables are, respectively, reported gross income, taxable income and net tax. In column 4 the audited taxpayer sample is restricted to taxpayers audited in 2012. In column 5 the audited taxpayer sample is restricted to taxpayers audited after 2008. In column 6 estimates are obtained applying CEM weights. In column 7 estimates are obtained applying CEM weights. In column 8 estimates are obtained considering the whole sample of audited and non-audited taxpayers (i.e., without matching). In column 9 estimates are obtained using income deciles instead of quartiles as matching variable. Robust standard errors are in parentheses, clustered at the taxpayer level. Individual controls are a dummy for “coherent” or “congruent” taxpayer and a dummy for “non-coherent” and “non-congruent” taxpayer (the omitted category includes taxpayers for which the presumptive revenue is not defined and for which, accordingly, the “coherent” and “congruent” variables have missing values)

***Significant at 1% level

As a further check of the robustness of our results, we replicate the analysis on the subsample of taxpayers audited in 2012, for which we observe a complete auditing cycle and we are able to conduct both parallel trends analysis and placebo regressions (see Sect. 4 and Table 3). Estimates in column 4 show that the audit effect is still positive and significant. The point estimate of β_1 is lower than when we use the whole sample of audited taxpayers. This difference is expected because in this case the post-audit period is just one year, and we have shown before that the audit's effect is larger in the second and third year after audit. In column 5 we restrict the sample to taxpayers audited after 2009, then excluding audited taxpayers for which we do not observe a change in the treatment status and that in the fixed effects model are included as controls. We confirm a positive effect of audit on subsequent tax compliance that is pretty similar in size to the one we find on the full sample of audited taxpayers.

When we use the logarithm of reported income as dependent variable (and restrict the sample to non-negative income) we find that audits are related to a 9.7% increase in reported income (column 6), similar to what we find using income levels. The positive audit's effect is confirmed when we estimate our baseline regression using CEM weights (column 7) albeit it is somewhat smaller. In the two final columns of Table 7, we replicate estimation without matching and using a different set of matching variables, namely using deciles instead of quartile of baseline reported income. Although ex-ante matching should help guarantee similarity between treated and control samples, we obtain similar results when we omit matching and use the full sample of taxpayers (the effect size is just marginally lower in this case). Finally, results are unchanged (with a just marginally lower effect size) when we use deciles instead of quartile of pre-treatment reported income as matching variables.

7 A back-of-the-envelope calculation of the net tax-revenue effect of an audit

The net tax-revenue effect of an audit is the difference between its benefits and costs. Benefits are the sum of the additional tax yield (the direct effect) and the deterrence effect. They differ across audit outcomes, however. A null outcome-audit does not generate any direct effect, while audits with a positive adjustment generate different direct effects depending on the specific audit outcome (no taxpayer's reaction, settlement, legal dispute). More specifically, audits concluded with no taxpayer's reaction yield an amount of additional taxes lower than the initial positive adjustment because in some cases the tax debt is impossible to collect. According to IMF (2016) estimates, the rate of effective collection of a euro assessed is 41%.³⁰ The direct effect of audits concluded with a settlement yield an amount of additional taxes corresponding to the initial adjustment less the abatement. The direct effect

³⁰ The remaining 59% of the tax debt is not collected for instance because the debtor is insolvent or because other provisions protecting taxpayer's assets are applicable. For more details see IMF (2016).

of audits concluded with a legal dispute is lower than the initial positive adjustment for two reasons. First, it is possible that the dispute is completely or partially lost by the AE, which estimates the probability of this event equal to 35%. Second, as said, the rate of effective collection is 41%. Thus, the direct effect is at least equal to 26.7% (0.41×0.65) of the initial positive adjustment. Also the deterrence effect varies across audit outcomes. In terms of additional taxes, it can be computed by applying a hypothetical 27% average effective tax rate to the additional reported income for every audit outcome.³¹

The audit's benefits must be weighted against its costs. According to the OECD classification (OECD, 2015), these latter can be related to: (i) audit and other verification activities; (ii) enforced debt collection and (iii) dispute and appeals. To estimate these costs, we use confidential data provided by the AE.³² When the targets of audits are small businesses, every hour of activity has a cost of approximately 55 euro. Of these, 35 euro represent a direct cost (i.e., the hourly wage of a representative taxman) and 20 an indirect cost (i.e., the share of administrative costs attributable to the audit activity). The AE estimates that an audit on a small business requires 35 h of work. As a result, the overall cost of an audit is 1925 euro.

For every euro spent on audits, the AE estimates a cost of 8 cents for debt collection and a cost of 23 cents for disputes and appeals. Accordingly, the debt-collection average cost is 154 euro (0.08×1925) and the dispute and appeals cost component is 443 euro (0.23×1925). These latter costs are not borne in the case of a settlement because the debt is immediately paid. In the case of no taxpayer's reaction, only the cost of debt collection is borne, while all three types of cost have to be paid when audits end up with a legal dispute. Table 8 shows the different components of the net tax-revenue of an audit. The average net tax-revenue weighted by shares of audits by outcome type is around 3.8 thousand euro.

On the one hand, these are conservative estimates because they ignore the spillover effects of tax audits, which can be important in the Italian setting (Galbiati & Zanella, 2012), as well as the effect of multiple audits.³³ On the other hand, the net tax-revenue effect is different from the overall welfare effect. First, consumption of some private goods would have been financed by the (unpaid) taxes if audit had not taken place. Second, while most of the above-described administrative costs would

³¹ The additional reported income is equal to the difference between reported income by audited and non-audited taxpayers differentiated by outcome type, estimated using the same estimator and specification as in Table 4, column 4. The differences amount to 2429 euro in case of no taxpayer's reaction, 2079 euro in case of settlement, 2459 in case of legal dispute and -267 euro in case of null outcome. We prefer to multiply these differences by the hypothetical average rate of 27% rather than using the net tax since the latter is likely influenced by heterogeneities in individual elements (tax deductions, allowances, etc.) which would alter the comparison.

³² OECD uses the same information for maintenance of the Tax Administration Database.

³³ Auditing activity presents scale economies: the average cost of audits is likely to decrease with the number of audits on the same taxpayer because during the subsequent audits some fixed costs, such as those related to acquiring information on time-invariant taxpayer characteristics, have already been borne by the tax authority. Scale economies arise also when more tax returns are checked upon the same audit due to the fixed cost related to opening an audit. Thus, the average audit cost and the average audit cost per tax return are lower in the case of multiple audits.

cancel out since they correspond to wages paid to public employees, compliance has private costs whose magnitude is generally considered higher than the administrative costs of audits (Slemrod and Gilltzer, 2014).

8 Conclusions

During the last decades of the twentieth century, the theory of tax evasion was dominated by the Allingham-Sandmo model, where expected utility maximizing taxpayers decide the level of income to report considering the private costs and benefits of evasion. Subsequently, this model was criticized because, given the actual levels of sanctions and the frequency of audits, it predicts evasion much higher than that observed. This induced many scholars to look for alternative explanations of tax evasion more related to “intrinsic motivations” (Andreoni et al., 1998; Luttmer and Singhal, 2014).

More recently, some studies have highlighted that when detection probability is correctly computed taking the presence of third-party information into account, taxpayers’ behavior is more in line with the Allingham-Sandmo model (Kleven et al., 2011). The explanation of high tax compliance may also depend on taxpayers’ mistakes in estimating detection probability and penalties (Chetty, 2009). Overall, the “cynical” view of the taxpayer maximizing her expected gain from the tax evasion lottery has thus regained attention and credit in this literature.

Our paper contributes to this stream of research by studying the impact of audits on subsequent tax compliance for a large panel of Italian individual self-employed taxpayers using real-world operational audits. Both our econometric strategy and the Italian institutional setting allow us to address potential endogeneity related to non-random selection of taxpayers to be audited.

In line with the theoretical predictions of the expected utility model of tax compliance, we find a positive average effect of audits on reported income of self-employed workers of approximately 8.4%. However, when the taxpayer is found compliant we find no effect of audits on subsequent tax compliance. Using data on audit’s cost and our results, we estimate a net tax-revenue effect of audits of around 3.8 thousand euro.

Overall, the issue of the external validity of our results naturally arises because the deterrence effect is always related to the tax system, and citizenries differ among themselves with respect to the magnitude and nature of noncompliance, to the norms that matter, and to the institutional environment (Slemrod, 2016). Nevertheless, we believe our result is of general interest because the population we look at, namely self-employed, are not subject to 3rd party reporting in all tax systems.

Our analysis can be sharpened on some dimensions. First, with the current database, we were not able to estimate the private compliance costs borne by taxpayers both during and after audits, and to provide a meaningful analysis of the welfare impact of audits. Although some components of private costs are transfers irrelevant

Table 8 Audit net tax revenues

	Null outcome	No reaction	Settlement	Legal dispute	Average
Assessment (taxes)	0	12,023	8688	25,450	11,540
BENEFITS (1)					
Tax yield (Direct effect) (2)	0	4929*	8688.3316	6782**	5100
Deterrence effect ^o (3)	-24	891	668	861	599
Total (1)=(2)+(3)	-24	5820	9356	7643	5699
COSTS (4)					
Audit ⁺ (5)	1925	1925	1925	1925	1925
Debt collection ⁺⁺ (6)	0	154	0	154	77
Dispute and appeals ⁺⁺⁺ (7)	0	0	0	443	111
Total (4)=(5)+(6)+(7)	1925	2079	1925	2522	2113
NET TAX REVENUE EFFECT (8)=(1)-(4)	-1949	3741	7431	5121	3806
Share of total revenues (9)	0.11	0.19	0.54	0.16	
WEIGHTED NET EFFECT=(9)×(8)	-210	713	4036	812	5352

*0.41 of positive assessment; **0.41×0.65 of positive assessment; ^o0.27×audit effect (see footnote 20); ⁺the values correspond to 55 (hourly cost of audits)×35 (nr of hours required); ⁺⁺the debt collection cost (154=1925×0.08) amounts to 0.08 euro for each euro spent on audit; ⁺⁺⁺the dispute and appeals cost (443=1925×0.23) amounts to 0.23 for each euro spent on audit

to welfare (e.g., payments to consultants and lawyers), other components imply a net loss (e.g., the opportunity cost of time spent on avoiding and responding to audits). Additional data on the length of each audit, coupled with an estimate of the opportunity cost of time along the lines provided by the Doing Business Database of the World Bank, would allow the estimation of the private compliance costs.

Second, in this study we ignored audits regarding taxpayers audited more than once and on more than one tax return (i.e., multiple audits). A longer panel and more information on the process driving the selection of taxpayers for multiple audits would allow us to provide some evidence on how compliance responds to different intensity of treatment. When a second audit is carried out a few years after the first, we expect a larger effect of later audits because these latter could induce taxpayers to reinforce their belief of being targeted by the tax authority, leading to an additional upward revision of their subjective audit probability and, accordingly, to a further increase in reported post-audit income. A similar mechanism is likely in place in the case of taxpayers audited on more than one tax return. The analysis of the effect of multiple audits would be relevant for the estimate of the tax-revenues effect of audits as well.

Third, in this paper we have shown that the impact of audits tends to decrease along the reported income distribution, which suggests that enforcement policies may reduce reported income dispersion. In view of this result, looking at the distributional implications of enforcement policies is another potential extension of our study and, more in general, of this stream of research (Slemrod, 2016). Obviously, this would require the identification of the reasons why high reported-income taxpayers tend to respond less to audits.

Finally, looking at audit's effect on the different income components (i.e., reported costs separately from reported revenues) would be interesting to test whether costs move in the same direction as revenues in response to an enforcement initiative, thus reducing the response of reported taxable income. In this respect, recent evidence has shown that when firms are notified about discrepancies between their declared revenues and revenues reports from third-party sources, they increase reported revenues but offset almost the entire adjustment with increases in reported costs, resulting in only minor increases in total tax collection (Carrillo et al., 2014).

9 Availability of data and materials and code availability

Our empirical analysis is based on proprietary and confidential data. Hence, we cannot make these data available but, other than providing the program files and other details of the computations sufficient to permit replication, we are of course available to fully cooperate with investigators seeking to conduct a replication.

Appendix 1

See Table 9.

Table 9 Audit effect by reported income decile and quartile

	2007 deciles	2008 deciles	2007 quartiles	2008 quartiles
<i>Decile/quartile</i>				
I	13,419.755*** (447.515)	13,879.656*** (517.824)	9313.250*** (237.994)	9418.661*** (249.299)
II	7523.424*** (321.728)	7433.468*** (294.778)	4409.560*** (231.839)	4016.469*** (219.033)
III	5462.916*** (296.723)	5481.453*** (292.051)	2085.500*** (301.350)	1897.004*** (297.569)
IV	4066.892*** (356.783)	3916.024*** (350.892)	-6674.184*** (559.719)	-6167.586*** (562.685)
V	4137.382*** (405.522)	3464.594*** (368.332)		
VI	2720.893*** (419.782)	3210.430*** (430.383)		
VII	2393.175*** (477.450)	1655.542*** (467.910)		
VIII	781.670 (595.031)	140.595 (566.192)		
IX	-2030.331*** (706.245)	-2133.689*** (708.114)		
X	-15,197.375*** (1080.223)	-12,921.941*** (1082.810)		
Constant	-6803.051*** (1262.422)	-7182.033*** (1250.092)	-6912.028*** (1264.325)	-7230.205*** (1250.906)
Individual controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Observations	1,914,353	1,910,745	1,914,353	1,910,745
Number of id	383,897	382,709	383,897	382,709
R-squared	0.062	0.062	0.062	0.062

In columns 1 and 2 income deciles are based on, respectively, 2007 and 2008 reported income values. In columns 3 and 4 income quartiles are based on, respectively, 2007 and 2008 reported income values. Robust standard errors are in parentheses, clustered at the taxpayer level. Individual controls are a dummy for “coherent” or “congruent” taxpayer and a dummy for “non-coherent” and “non-congruent” taxpayer (the omitted category includes taxpayers for which the presumptive revenue is not defined and for which, accordingly, the “coherent” and “congruent” variables have missing values)

***Significant at 1% level

Appendix 2

See Table 10.

Table 10 Descriptive statistics and marginal tax rate by deciles of reported income

Reported income (decile)	Reported income				Marginal tax rate
	Mean	SD	Min	Max	
1	-3250.052	12,410.74	-222,958.8	0	0
2	5747.554	2858.287	1.085	9900.625	0.23
3	12,476.52	1359.923	9901.71	14,637.74	0.23
4	16,612.09	1151.085	14,638.82	18,611.01	0.23
5	20,738.33	1282.692	18,612.09	23,060.59	0.27
6	25,746.69	1617.889	23,061.68	28,654.85	0.27
7	32,259.19	2210.693	28,655.94	36,360.52	0.38
8	41,961.29	3516.356	36,361.61	48,655.74	0.38
9	59,373.26	7148.083	48,656.82	73,727.92	0.41
10	112,255.9	35,109.76	73,730.09	220,494.8	0.43

The source for the marginal tax rate is www.finanze.it

Acknowledgements This paper is part of a common research project between DEMS and the Italian Revenue Agency. We are particularly thankful to Dr. Stefano Pisani and to the Research Unit at the Italian Revenue Agency for providing us with the dataset and for their suggestions and clarifications on the nature and interpretation of the data. We wish to thank James Alm, Gareth Myles and Joel Slemrod for valuable comments and suggestions. The usual disclaimers apply.

Authors' contributions Not applicable.

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement. Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Advani, A., Elming, W., & Shaw, J. (2019). *The dynamic effects of tax audits*. Working Paper 414, May 2019, CAGE, University of Warwick.
- Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3–4), 323–338.
- Andreoni, J., Erard, B., & Feinstein, J. (1998). Tax compliance. *Journal of Economic Literature*, 36(2), 818–860.

- Best, M. C. (2014). The role of firms in workers' earnings responses to taxes: Evidence from Pakistan. *mimeo*.
- Cabral, A. C. G., Kotsogiannis, C., & Myles, G. (2014). *Self-employment underreporting in Great Britain: Who and how much?* Tax Administration Research Centre. University of Exeter Discussion Paper no. 010-14.
- Carrillo, P., Pomeranz, D., & Singhal, M. (2014). *Dodging the taxman: Firm misreporting and limits to tax enforcement*. Harvard Business School Working Paper n. 15-026. Forthcoming in *American Economic Journal: Applied Economics*.
- Chetty, R. (2009). Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance. *American Economic Journal: Economic Policy*, 1(2), 31–52.
- D'Agosto, E., Manzo, M., Pisani, S., & d'Arcangelo, F. M. (2018). The effect of audit activity on tax declaration: Evidence on small businesses in Italy. *Public Finance Review*, 46(1), 29–57.
- DeBacker, J., Heim, B. T., Tran, A., & Yuskavage, A. (2015). Legal enforcement and corporate behavior: An analysis of tax aggressiveness after an audit. *The Journal of Law and Economics*, 58(2), 291–324.
- DeBacker, J., Heim, B. T., Tran, A., & Yuskavage, A. (2018). Once bitten, twice shy? The lasting impact of IRS audits on individual tax reporting. *The Journal of Law and Economics*, 61(1), 1–35.
- Galbiati, R., & Zanella, G. (2012). The tax evasion social multiplier: Evidence from Italy. *Journal of Public Economics*, 96(5), 485–494.
- Gemmell, N., & Ratto, M. (2012). Behavioral responses to taxpayer audits: Evidence from random taxpayer inquiries. *National Tax Journal*, 65(1), 35–58.
- Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20, 1–24.
- IMF. (2016). *Italy. Technical assistance report-enhancing governance and effectiveness of the fiscal agencies*. IMF Country Report N.16/241.
- Italian Ministry of Economy and Finance. (2016). *Relazione sull'economia non osservata e sull'evasione fiscale e contributiva*. Available at: http://www.mef.gov.it/inevidenza/documenti/Relazione_evasione_fiscale_e_contributiva_0926_ore1300_xversione_definitivax-29_settembre_2016.pdf
- Kleven, H. J. (2014). How can Scandinavians tax so much? *The Journal of Economic Perspectives*, 28(4), 77–98.
- Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651–692.
- Løyland, K., Raaum, O., Torsvik, G., & Øvrum, A. (2019). *Compliance effects of risk-based tax audits*. CESifo Working Paper No. 7616.
- Luttmer, E. F. P., & Singhal, M. (2014). Tax morale. *The Journal of Economic Perspectives*, 28(4), 149–168.
- Muralidharan, K., & Prakash, N. (2013). *Cycling to school: Increasing secondary school enrollment for girls in India*. IZA Discussion Paper No. 7585.
- Oecd. (2015). *Tax administration 2015. Comparative information about Oecd and other advanced and emerging economies*. Oecd Publishing.
- Paulus, A. (2015). *Tax evasion and measurement error: An econometric analysis of survey data linked with tax records*. Institute for Social and Economic Research, University of Essex Working Paper No. 2015-10.
- Pissarides, C. A., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, 39(1), 17–32.
- Santoro, A., & Fiorio, C. V. (2011). Taxpayer behavior when audit rules are known: Evidence from Italy. *Public Finance Review*, 39(1), 103–123.
- Slemrod, J. (2004). The economics of corporate tax selfishness. *National Tax Journal*, 57(4), 877–899.
- Slemrod, J. (2016). *Tax compliance and enforcement: New research and its policy implications*. Ross School of Business Working Paper 1302. University of Michigan.
- Slemrod, J., Blumenthal, M., & Christian, C. (2001). Taxpayers response to an increased probability of audit: Evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, 79, 455–483.
- Slemrod, J., & Gillitzer, C. (2014). Editor's choice insights from a tax-systems perspective. *Cesifo Economic Studies*, 60(1), 1–31.
- Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1–2), 305–353.