Behavioral Responses to Tax Audits: Between Deterrence and Approval[*]

Shafik Hebous[¤], Zhiyang Jia[§], Knut Løyland[†], Thor O. Thoresen[‡], and Arnstein Øvrum[**]

Preliminary, 20 February 2019

Abstract

Auditing the income tax return not only ensures that the correct amount is paid in the year of the intervention, it is expected to affect the tax compliance in the subsequent years too. A random audit selection scheme operated by the Norwegian tax administration is used to identify magnitude and duration of post-audit deterrence effects. Moreover, we explore to what extent there is a counteracting "approval effect" too, among the taxpayers found to be compliant by the audit. We find estimates in accordance with a modest deterrence effect, however statistically significant only in the first year after the audit. Behind this we see substantially larger effects in the non-compliant group, lasting for five years after the audit. The compliant taxpayers, however, show no signs of behavioral adjustments.

Keywords: Tax compliance, deterrence of audits, administrative data

JEL codes: H26, C23

[*]This research has been carried out as a collaboration between the Norwegian Tax Administration and Oslo Fiscal Studies (Department of Economics, University of Oslo). We thank the Norwegian Research Council for the support of Oslo Fiscal Studies.

[¤]International Monetary Fund and Oslo Fiscal Studies (University of Oslo), email: Shebous@imf.org

[§]Statistics Norway, email: zhiyang.jia@ssb.no

[†]Norwegian Tax Administration, email: knut.loyland@skatteetaten.no

[‡]Statistics Norway and Oslo Fiscal Studies (University of Oslo), email: thor.olav.thoresen@ssb.no

[**]Norwegian Tax Administration, email: arnstein.ovrum@skatteetaten.no

1. Introduction

Tax audits hold the promise of enhancing tax compliance. Although effects of audits on tax compliance have been extensively studied, see Andreoni, Erard and Feinstein (1998), Slemrod (2007), and Alm (2019), there is less information about how effects develop over time. From a perspective of efficient tax administration, it is crucial to obtain information on to what extent audits deter taxpayers from non-compliance also in the years after being undergoing inspection. As we have access to an administrative dataset (of the Norwegian Tax administration) following a large number of individual taxpayers before and after being audited, this analysis adds to the relatively scarce literature that discusses effects of audits in a longer time perspective.

A clear advantage of the present study is that data are generated by a purely random audit selection procedure. The schedule is as follows. Most Norwegian taxpayers receive a complete prefilled income tax return based on an extensive third-party reporting system. This means that many taxpayers do not need to make amendments, just confirm that they accept.[1] However, although wide-ranging, not all items are third party reported, and taxpayers have scope for entering information missing from the form. With respect to deductions, there is an item on the income tax return form named "Other deductions", used by both wage earners and self-employed for claiming deductions not already recorded. This particular item is subject to further inspection by the tax administration, which has generated the data set of the present study. Critically, and as already noted, the selection process is purely random. Administrative secrecy prevents us from revealing the precise criteria, but the principles suffice for the understanding of the procedure: XX percent of taxpayers with deductions under "Other deductions" in the range from YY,000 to ZZ,000 Norwegian kroner (NOK) are assigned to further control. The data set generated by these audits includes information on all taxpayers eligible for the inspection, not only the selected, and all are observed in up to six years after the control.

The present analysis uses data for approximately 30,000 individuals from these administrative inspections in 2009, 2010 and 2011, adopting a quasi-experimental empirical design by letting the year of the audit represent a demarcation line. This means that we can compare deduction behavior before and after the audit for the treated (audited) and the control group (the non-audited). Thus, the treatment refers to the attention given to taxpayers in terms of being notified and informed that their deduction claim has been further investigated. Obviously, and as we soon will return to, there are two groups among the audited, the "compliant" and the "non-compliant", where the latter group consists of taxpayers whose income tax return have been examined and dismissed due to irregularities on the item "Other deductions". These taxpayers may also get fined, but, in practice, because of the magnitude of the criminal act, they are not. Thus, the intervention from the tax authorities comes in the form of "attention" more than inflicting penalties. In the "compliant group" we find those who have been

[1] In fact, if the person (for some reason) does not review it, it is regarded as filed and accepted. The extensive third-party reporting scheme means that information goes directly from, for example, employers (wage income), banks (interest income and wealth), and charitable organizations (donations are deductible in income).

checked and cleared. Importantly, most compliant taxpayers are also informed about the inspection, as they normally are asked about further documentation of their claims.[2]

Firstly, this study presents estimates of the average effect of audits, with estimates derived from a simple treatment-control/before-after empirical design, utilizing the panel dimension of the data.[3] Although the procedure opens up for accounting for individual fixed effects, we do not expect such controls to influence results. It is the random selection of taxpayers into audit that defines the key to obtain non-biased estimates.

It follows that an investigation along these lines holds the promise of contributing to the understanding of one of the most central question of the literature on effects of auditing, namely to what extent attention by the taxman deters or encourages non-compliance in the years after inspection. The informational content of an audit can be referred to a type of Bayesian updating, see Snow and Warren (2007), i.e., the experience leads to a revision of beliefs. The literature refers to both a "target effect" (Hashimzade, Myles and Tran-Nam, 2013), which means that audits deter under-reporting because the agents perceive that chances for another inspection is high, and "a bomb-crater" effect (Maciejovsky, Kirchler and Schwarzenberger 2007; Mittone, Panebianco and Santoro, 2017), a "bomb" would not strike exactly the same place again.

Further, the auditing itself reveals information about the treated, in that it divides them into compliant and non-compliant taxpayers, which we shall utilize to further add to the understanding of how people react to the attention from the tax authorities. Thus, in the second part of the empirical investigation we discuss the behavioral responses to audits by expanding on separate effects of "negative" and "positive" attention. Although audits may discourage future illegal activities for taxpayers who have been caught evading, the experience of being checked and cleared may generate other reactions. Given that we, as already explained, apply a difference-in-differences identification strategy, the key to obtain credible estimates of subgroup behavior is that the deduction trend prior to the audit in both groups is close to what we observe in the non-audited group.

With respect to the second part of our study, the focus on subgroups of the treated, two studies are very close to the empirical design of the present study, as they address the behavior of both compliants and non-compliants – Gemmell and Ratto (2012) and Beer et al. (2015). Both studies find that the non-compliant taxpayers increase their subsequent compliance after an audit. But notably, both analyses find evidence suggesting that the compliant taxpayers react in the opposite way, as they become less compliant after being audited.

The paper is organized as follows. In Section 2 we survey the literature on responses to audits, whereas Section 3 describes in more detail the institutional setting that has produced the audit data

[2] We are aware that there are some taxpayers in the compliant group who are not informed about their selection into auditing. As we shall return to, these taxpayers essentially behave like taxpayers in the non-treated group and therefore represents a bias towards the mean.
[3] As a matter of terminology, note that the "treatment" of the present study is a "control".

exploited here. In Section 4 we present the empirical framework and estimation results for the average overall effect of audit, whereas Section 5 presents regressions results when the audited are divided into compliant and non-compliant taxpayers. Section 6 shows results of several robustness tests and Section 7 concludes.

2. Responses to attention from audits

Audits influence the tax collection directly, as it means that additional revenue is collected from people not abiding by the rules. Here, the attention is on the effects of audits in a longer time perspective – to what extent the taxpayers are deferred by the tax audit, and thereby change their subsequent compliance behavior, often referred as the indirect effects of audits.[4] As the previous literature finds that the indirect effects outweigh the direct effects (Dubin, 2007; Ratto, Thomas and Ulph, 2013), from an efficiency of tax administration point of view it is imperative to enhance the knowledge about magnitudes.

In the Allingham-Sandmo model (Allingham and Sandmo, 1972), as summarized in Sandmo (2005), we have that net income, *Y*, of the taxpayer is defined by $Y = W - t(W - E)$, where *W* is gross income, *E* is the amount of underreporting, and *t* is the proportional tax rate. Given that there is a penalty paid on the evaded tax (Yitzhaki, 1974), the net income, *Z*, can be seen as $Z = (1-t)W - (\theta - 1)tE$, where $\theta(>1)$ is the penalty rate. As the taxpayer's subjective probability of detection is *p*, and he maximizes $V = (1-p)U(Y) + pU(Z)$, the first order condition becomes $\frac{U'(Z)}{U'(Y)} = \frac{1-p}{p(\theta-1)}$.[5] The main subject of the present analysis is to discuss how the amount of underreporting, *E*, is determined in Period 2, when the taxpayers have received attention in Period 1. As we soon shall return to, there are no penalties in our case, so the main line of reasoning builds on the attention of the tax administration influences the perception of *p* in Period 2. Snow and Warren (2007) refer to this process as Bayesian updating, since the detection probabilities are updated based on the experience from Period 1.

Given this theoretical set-up, there are two important characteristics that represent deviations from the reasoning. Firstly, in general, the tax administration will not issue fines to the non-compliant taxpayers, since the gravity of the criminal act must be characterized as modest. However, as it is generally acknowledged tax evasion involves more than amoral cost-benefit calculations of the agents,

[4] This effect may also be characterized as "corrective" (Gemmell and Ratto, 2012) or a "direct deterrent effect" (Alm, Jackson and McKee, 2009). The "indirect deterrent effect" then refers to spillover effects on the non-audited, see discussion of spillover effects in for example Fortin, Lacroix and Villeval (2007).

[5] It should be noted that this model confronted with values of fines and audit probabilities, in most cases, overshoots the level of tax evasion observed. It is a general understanding that other explanations must be added to the framework (Hashimzade, Myles and Tran-Nam, 2013).

we believe that there are other (intrinsic) motivations for being compliant. The basic model has been extended in several directions, including accounting for moral sentiments of guilt and shame (Erard and Feinstein 1994) and social conformity effects (Myles and Naylor 1996; Fortin, Lacroix, and Villeval 2007). Thus, we believe that the mechanism of deterrence (if there is one) works even without issuing actual fines. Secondly, a key characteristic of the updating of the present study is that we have two types of taxpayers in Period 2: taxpayers who in Period 1 have been found to comply and not to comply, respectively. As soon shall return to, these two groups are not expected to show the same response to the attention given to them from the tax administration.

The literature is dominated by studies which do not separate between compliant and non-compliant taxpayers. With respect to the behavior of the audited in general, there are mixed results on effects of post-audit behavior (Alm, 2019). Although overview studies report that there is a tendency to reduce tax evasion, compatible with deterrence, effects are small or even negligible (Andreoni et al., 1998; Gangl et al., 2014). One potential reason is that tax enforcement may crowd out the intrinsic motivation of paying taxes (Luttmer and Singhal, 2014; Dwenger et al., 2015; Mendoza, Wielhouver and Kirchler, 2017). However, Slemrod, Blumenthal and Christian (2001) find evidence in accordance with the most intuitive mechanism – a random sample of taxpayers report higher income after being warned about future close examination of their income returns. This can be characterized as a target effect (Hashimzade, Myles and Tran-Nam, 2013), simply stating that agents perceive that an audit is likely to be followed by another in the future. Along the line of Tversky and Kahneman (1974), a possible explanation is that individuals' increase their awareness of audits, the decision-makers are inclined to believe that an event is more likely to happen if the event (attention of the tax administration) is easily retrieved from the memory. In contrast, some studies refer to a "bomb-crater effect" (Maciejovsky, Kirchler and Schwarzenberger, 2007; Mittone, Panebianco and Santoro, 2017), referring to individuals perceiving that the risk of being audited again falls immediately after an audit, because a bomb would not strike exactly the same place again (hence soldiers in WW1 hide in bomb craters). This can be explained by misperception of chance – in this case it is expected that a random event (audit) is less likely to happen again if it just happened.

With respect to the distinction between compliant and non-compliant taxpayers among the audited, the post-audit behavior of the compliant may be particularly hard to predict, see Beer et al. (2015). One may reasonably expect that the taxpayers in this group just simply continue their deduction behavior, as before. However, there are reasons for finding more illegal behavior in this group after audit. For example, the previously compliant taxpayers now infer that the risk of future checks is low and thereby exploit the information provided by the tax administration to decrease the subsequent tax burden. Further, as already discussed, given that it is widely established that taypayers are motivated by intrinsic motivation, auditing level may signal distrust in taxpayers and lead to the perception that the tax authority and its enforcement actions are excessive and unfair, which in turn lead to reduced compliance. Also, if the compliant in our data in reality are non-detected non-

compliant taxpayers, they may become even more motivated to continue their illegal activities by the lack of detection.

Effects of audits in a longer time perspective are both discussed in studies based on laboratory experiments and analyses of actual data. With respect to the former type of studies, Alm, Jackson and McKee (2009) use laboratory experiments to examine the compliance impact of types of information dissemination regarding audit frequency and find results conditional on whether the taxpayer is well informed about the audit rate prior to filing. Results suggest that it would be advantageous to pre-announce audit rates. Turning to studies that discuss compliance behavior by analysis of tax data, both Dubin, Graetz and Wilde (1990) and Dubin (2007) show strong positive effects on compliance after the IRS (Internal Revenue Service) has shown interest in the filing of taxpayers. Newer studies confirm that audits or other types of interventions seem to deter taxpayers from non-compliance.[6] However, the results of both Advani, Elming and Shaw (2017) and DeBacker et al. (2017) point to relatively moderate effects. Like as in Kleven et al. (2011), both studies report that effects are stronger with respect to self-reported income components than third-party reported income. Moreover, both analyses find short-lived effect; for example, in Advani et al. the effects on self-employment income and dividends die away after about four years.

There are two empirical investigations that are particularly close to the present study in that they distinguish between effects of compliant and non-compliant taxpayers – Gemmell and Ratto (2012) and Beer et al. (2015). Gemmell and Ratto investigate the response of U.K. of a sample of taxpayers to randomly selected to audits in the year 2000, consisting of both business owners and "personal" taxpayers. If the direct yield (additional tax plus penalties) is positive, the taxpayer is classified as non-compliant, whereas no change defines compliance. The responses are measured in terms of comparing declared tax in three years before the audit to three years after the audit, using difference-in-differences regression analysis to identify effects. Whereas all the involved are informed about the audits, in some cases the closure of the inquiry is well into the post-audit observation period, which is found to influence results. Moreover, and important with respect to the present study, two specifications are estimated: one measuring the overall effect on the audited, and another allowing for separate effects of the non-compliant and the compliant taxpayers. In the latter case, the effects of the two groups are estimated jointly, which implies that the same group of taxpayers (the non-audited) are used as control group for the two treatment groups.[7] Controlling for individual fixed effects, observable and unobservable characteristics independent of time, is claimed to account for pre-audit differences between the three groups. Results suggest that taxpayers are deferred by audits, as audited taxpayers who were found to be compliant reduced their subsequent compliance. But the opposite response was observed for taxpayers previously found to be noncompliant. Thus, the authors write (p.

[6] However, DeBacker et al. (2015) find that U.S. firms reduce tax payment immediately after an audit and then increase it gradually to the pre-audit level.
[7] As we soon shall return, this empirical strategy is also followed here.

55): “These results serve to highlight the importance of testing for the responses of the so-called compliant and noncompliant subgroups separately to avoid conflating their different responses.”

Interestingly, Beer et al. (2015) find that effects of audits differ with respect to compliance and non-compliance in the same manner as reported by Gemmell and Ratto (2012). Beer et al. use the impact of enforcement activity and subsequent compliance income reporting of sole proprietors in the U.S. in their analysis, comparing a random sample of filers who were audited after filing in 2007 to a control sample of non-audited under the same schedule (Schedule C filers). Thus, this means that the selection of audited taxpayers is operational rather than random, which is important when interpreting the magnitudes of the effects. The identification strategy is close to the approach of Gemmell and Ratto (2012), in that they employ a difference-in-difference empirical design. Three years after the audit a positive reporting effect is seen among the non-compliant taxpayers, in accordance with a deterrent effect, whereas audits have a detrimental impact on the reporting behavior of taxpayers who do not experience an additional tax assessment – see above for possible explanations. Moreover, the sensitivity tests include both specifications with selection on both observables and unobservables, employing matching techniques and the selection model of Heckman (1978), respectively.

3. Third-party reporting, random audits and filters

Before we present the empirical approach and results, we shall briefly refer to the institutional setting from which the data are derived. The Norwegian third-party information schedule has been developed to the extent that most taxpayers are not in contact with the tax administration before receiving the form for approval: most income tax returns are prefilled based on third-party information from employers, the financial sector and others. For example, as donations are deductible up to a threshold in the taxation of ordinary income in Norway, the recipients of donations (as the Red Cross) report the individual donations directly to the tax authorities, and the information is in turn assigned the individuals. The taxpayer is therefore usually presented to a complete prefilled income tax return and is asked to approve before filing. Of course, if he finds errors or incomes/deductions not reported, he makes amendments. As the Norwegian tax administration is well underway to make the tax system fully electronic, the income tax return is typically electronically filed, and, accordingly, most people make amendments by addressing their income tax return directly through the internet. Thus, there is usually no administrative staff of the tax authorities involved in this.

In the present analysis we focus on an item on the income tax return which is frequently used to report additional deductions, the “Other deductions”, used for claiming deductions not already recorded through other items of the third-party reporting system. For example, medical costs or costs from receiving substantial help from a lawyer, to be able to earn an income in the labor market, can be deducted here. A problem from a tax administrative point of view is that any taxpayer can make

substantiated or unsubstantiated claims, in the latter case using this item to illegally reduce the tax burden. Thus, the claimants are exposed to audits.

In general, the tax administration employs a whole range of filters to select individuals for audits. With respect to this particular item, the selection is based on establishing a pool of taxpayers who have claimed "Other deductions" in the range from YY,000 to ZZ,000 NOK. From this sample, a subsample of XX percent is randomly selected for further investigation; administrative secrecy prevents us from revealing the precise criteria. In the present study we utilize information from this type of auditing for the years 2009, 2010 and 2011.[8]

From an empirical identification perspective, it is important to note that not only the non-compliant are informed about the attention form the tax administration, but also most of the compliant taxpayers.[9] Of course, the non-compliant taxpayers are informed since the tax authorities adjust their income tax return.[10] Some of the compliant taxpayers may go through the process without notification, if they already have provided all the necessary documentation needed. However, most of them would have been asked to provide additional information, which implies that they are aware of the attention of the tax administration.

4. Average effect of audit

4.1. A simple difference-in-differences (DID) set-up

The key identifying feature of the analysis is that the audit is completely random. This means that identification of effects can be obtained by techniques associated with natural experiments. Since we have observation of taxpayers belonging to the pool of taxpayers in danger of being exposed to audit, independent of being controlled or not, and since we observe them both before and after the audit, we shall employ the difference-in-differences (DID) estimator in the following. We define a binary regressor, D_i, as

$$(1) \qquad D_i = \begin{cases} 1 \text{ if individual } i \text{ receives treatment} \\ 0 \text{ otherwise} \end{cases}.$$

Further, let the dependent variable, the deduction behavior, be symbolized by y_{0i} and y_{1i}, before and after the treatment, respectively. The effect of the audit can then be seen as an average treatment effect for the treated (ATT),

$$(2) \qquad E\left[y_{1i} \middle| D_i = 1 \right] - E\left[y_{0i} \middle| D_i = 1 \right] = E\left[y_{1i} - y_{0i} \middle| D_i = 1 \right].$$

[8] After 2011 the tax administration changed the audit procedures.
[9] Obviously, effects may depend on how much information that is conveyed to the agents. For example, taxpayers in the U.S and Denmark, randomly selected for audits, are informed about the investigation being based on random selection.
[10] In some grave cases the taxpayers may be fined too, but such incidences would be rare given the amounts involved here.

This simple framework relies on the behavior of the non-audited representing the counterfactual of the audited, an assumption which we believe is defensible by the random assignment of auditing in our case. Note that so far we have not distinguished between compliant and non-compliant members of the audited group; we shall return to this distinction in Section 5.

4.2. Data descriptions

In Table 1 we present the number of observations we have had available for this study, allocated on different groups. As the data are collected from three different audits, in 2009, 2010, and 2011, and the observation period starts in 2008 and lasts until 2015, we observe the taxpayers in minimum five years and in maximum eight years after the audit. The table shows that approximately 30,000 taxpayers qualified for audits in the three years, from which approximately 3,500 have been assigned to audits by the random selection. As the demarcation line is the year of the audit, the number of observations in each year varies: if the individual is audited in 2011, we observe he/she in three years before the audit and in four years after the audit, whereas a person audited in 2009 is observed in six post-audit years, but only in one year before the audit.

Table 1. Number of observations in samples

	Number of individuals in panel	Number of observations
Non-treated	26,775	197,396
Audited	3,476	22,646
Compliant	2,238	14,338
Non-compliant	1,238	8,308

Table 2 presents descriptive statistics for the two groups. As expected, given the random assignment, there are no clear differences in characteristics between the two groups. We note that the average claimed deduction is around 23,000 NOK in both groups.[11] After the audit, the tax authorities have decided to reduce the deduction by approximately 9,000 NOK on average among the audited taxpayers, due to non-verified claims.

[11] Use exchange rates for 2015 to convert to euros and US dollars: 1€=NOK 8.95 and 1$=NOK 8.07.

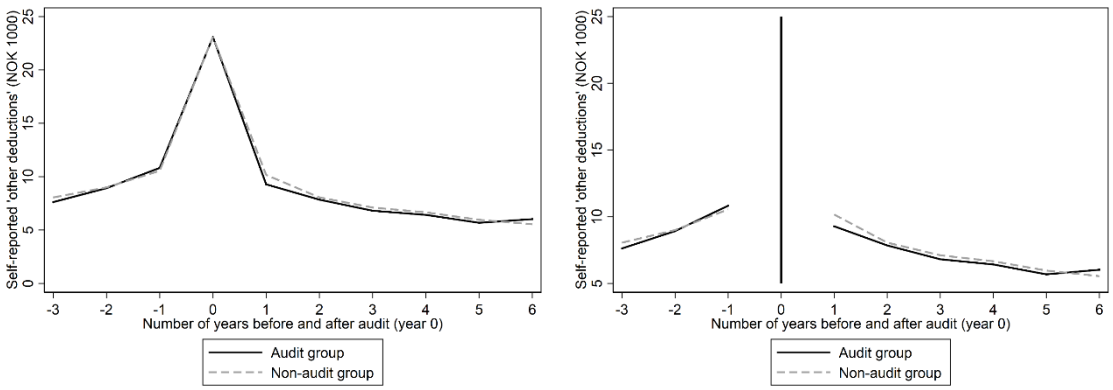
Table 2. Descriptive statistics, audited and non-audited

	Audited	Non-audited
Claimed "Other deductions"	23,055	23,101
	(11,502)	(11,290)
Direct correction in "Other deduction" due to audit	8,993	-
Self-employed	0.210	0.208
	(0.408)	(0.406)
Female	0.285	0.290
	(0.451)	(0.454)
Age	48.9	48.4
	(13.0)	(13.0)
Temporary work migrant	0.048	0.060
	(0.213)	(0.237)
Total deductions	216,280	216,652
	(209,417)	(205,798)
Third-party rep. deductions	143,635	141,245
	(113 261)	(108,897)
Total gross income	774,454	772,372
	(747,502)	(736,564)
Total third-party rep. gross income	608,880	607,771
	(483,927)	(484,428)
Observations	3,476	26,775

Note: Standard deviations in parentheses

Next, in Figure 1, the trend in the deduction behavior is shown. Recall that as the data are from the period 2008–2015 and that there were audits in 2009–2011 (year 0), we observe the taxpayers up to three years before the audit and a maximum of six years after. Note that the sample selection rule applied by the tax authority implies that only individuals whose deduction values lie in a given interval are selected into our analysis. Thus, the distribution of deduction values in year 0 is substantially different from the rest of the years, as clearly depicted by the left-hand diagram of Figure 1.

Figure 1. Mean deductions before and after audit, audited and non-audited



Note: Left-hand diagram includes deductions in the year of the audit, year 0. Right-hand diagram is representative for the observations used in the regressions, i.e., year 0 is excluded.

Including data from the year of the audit or not in the empirical analysis would not likely influence results, as this year represents a similar deduction outlier for both the audited and the non-audited. We have decided to leave out data of the year $t = 0$ in the empirical investigation, which

means that the right-hand diagram is representative for the data employed in the study. We note that the diagram shows strong support to the pre-treatment parallel trend assumption, an important identifying assumption of the DID identification strategy.

To further address to what extent there is individual persistence in the deduction behavior over time, we estimate a simple autoregressive model of order 1, AR(1). We obtain an estimate of the autocorrelation between year t and $t-1$ at the level of 0.006, with a standard error of 0.002. Although the autocorrelation estimate is statistically significantly different from zero, the magnitude is small. Hence, we ignore potential persistence in the deduction behavior in the following.

Behind the average measures reported in Figure 1, the distribution of deduction values is highly uneven. There is a large share (over 60%) of individuals who report zero deduction.

4.3. Estimates of the average treatment effect

Given the random assignment, one could simply compare the sample means of the audited and non-audited group to obtain an estimate of the average treatment effect. In the following, we apply a regression DID framework very close to the theoretical outline. The effect of audit on the deduction behavior, y_{it}, can be seen as

$$(3) \qquad y_{it} = \alpha + \delta_s \left(D_{it} \times k_s \right) + \lambda_t + \mu_i + \varepsilon_{it} \quad s \in \{1,2,...,6\},$$

where, as in Equation (1), the binary regressor, D_{it}, takes the value 1 for the audited. However, we estimate effects for each post-treatment year, indicated by k_s, which is a dummy variable taking the value 1 for each post-audit year, s, after the audit. This means that we estimate a sequence of parameters for the post-audit effects of audits, controlling for a constant, α, year-specific effects, λ_t, and the individual fixed effect, μ_i.

When we estimate the average effect of treatment, i.e., not providing estimates for each year after audit, we obtain an estimate of -750 NOK.[12] Thus, on average the audit resulted in a reduction in claimed deductions of 750 NOK. As a tentative estimate of the overall effect of this type of audit, one may use this estimate as a yearly average effect per year for the first six years after the audit, obtaining a measure for the increased tax payments resulting from it. Recall that approximately 3,500 taxpayers have been checked in the present study, see Table 1.[13]

[12] The estimate is statistically significant at the 0.1 level.

[13] Of course, an audit generates revenue for the year of the audit (the direct effect, see Table 2). The main contribution of the present study is to provide estimates of the after-audit effects. A very simple "back-of-the-envelope" calculation suggests that due to this particular audit (with filters as described in Section 3), tax revenue increased by 4.4 million NOK. This follows from the tax base (after deductions) being taxed by a flat rate of 28 percent: $(750 \times 6 \times 3500)0.28$. Of course, there could be spillover effects on other items, which should be taken into account; we will return to this. There are also possible "network effects" influencing the overall deterrence effect, obviously hard to measure.

Next, we open up for differential effects over the time period after the audit, as signified by Equation (3). Although all point estimates suggest that taxpayers reduce their reporting of deductions because of the audit, effects are only statistically significant[14] in the first year after the audit. Figure 1 clearly shows that the reduction behavior (for "Other deductions") returns to more normal levels after the audit year. It follows that the estimates pick up the small difference between the audited and the non-audited after year 0, as seen in Figure 1. In Section 5 we shall see that this average treatment effect is composed of differing behavioral changes from taxpayers who been found compliant and non-compliant through the audit.

Table 3. Effects of audit on post-audit deduction behavior

Year after audit	Coefficient	Estimate	*t*-value
First	δ_1	-1,272*** (460)	-2.76
Second	δ_2	-572 (454)	-1.26
Third	δ_3	-626 (460)	-1.36
Fourth	δ_4	-557 (465)	-1.20
Fifth	δ_5	-479 (482)	-0.99
Sixth	δ_6	-189 (592)	-0.32
Observations		177,161	

Notes: Fixed effect estimation based on panel data 2008–2015. Robust standard errors in parentheses.
$*p < 0.1, **p < 0.05, *** p < 0.01$

4.4. Distributional treatment effect

As the results so far show that there is a negative shift in the mean deduction after the audit, it is interesting to what extent the shape of the deduction distribution has been changed. Here we follow Hernæs and Jia (2013) and Brinch, Hernæs and Jia (2017) and look at the Complementary Cumulative Distribution Functions (CCDF) $\bar{F}(y \mid X) = \Pr(Y > y \mid X)$ before and after audit. To be more precise, we use a series of logit specifications to model the conditional complementary CDF for a number of values of y. This allows a simple application of the difference in difference technique to identify the treatment effect of the audit.

For any given value of $y \geq 0$, we assume that, for individual i over the observation period

$$\Pr(y_{it} > y_k) = G(\alpha_k + \delta_{sk}\left(D_{it} \times k_s\right) + \lambda_{tk} + X_{it}\beta_k + \varepsilon_{ikt}) \quad s \in \{1,2,...,6\}, \tag{4}$$

where X_{it} denotes individual characteristics and G represents the logit function. We estimate the logit specification 5 separate times, letting y_k vary from NOK 0 to NOK 100,000 by increments of NOK 5,000 such that $y_k \in \{0,5000,10000,...,100000\}$. The coefficients δ_{sk} for $s \in \{1,2,...,6\}$ measure the time-specific audit effects.

[14] Recall that due to our empirical design, there are less observation in the last years of the period, which likely contribute to the lack of statistical significance.

The graphical illustration in Figure 2 is based on five separate estimations, one for each of the deduction levels described above. For each estimation, we find the marginal effect of audit evaluated at the covariate value equal to the average of the treatment group. These marginal effects are equal to the difference in the post-audit and pre-audit probability of a deduction larger than a given level of y: $\Pr(y_{it} > y \mid D_{it} = 1, X_{it}) - \Pr(y_{it} > y \mid D_{it} = 0, X_{it})$.

Figure 2 shows the estimated marginal effects with 95% confidence envelops over different deduction levels for the year right after audit ($s = 1$). The effects from the other years are similar but much weaker. The figure shows that the audit affects deduction claims on both the intensive and extensive margin. There are fewer individuals who claim deduction after the audit, and effects on the intensive margin are uneven across deduction levels, with largest effect observed in the interval $[5000, 25000]$. The corresponding shifts in the probability of being in different intervals of the claimed deduction distribution are reported in Table 4.

Figure 2. Audit effects on the distribution of deductions, the year after audit

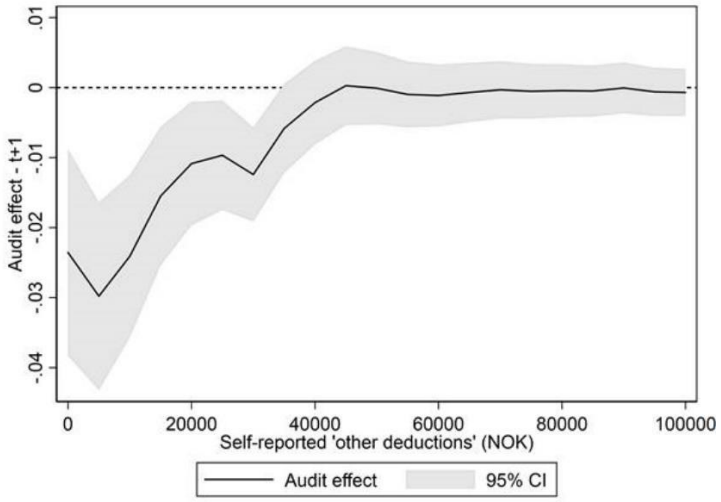


Table 4. Descriptive statistics for the distribution of deductions. Compliant, non-compliant and non-audited

	Audit effect	
Interval	Estimate	Standard error
No claiming (NOK 0)	0.024	0.008
NOK 0 – NOK 5,000	0.006	0.010
NOK 5,000 – NOK 25,000	-0.020	0.008
NOK 25,000 – NOK 40,000	-0.008	0.005
NOK 40,000 – NOK 50,000	-0.002	0.004
> NOK 50,000	-0.000	0.003

5. Distinguishing between compliant and non-compliant taxpayers

5.1. Extended information on the audited

The auditing process generates two distinctively different groups among the treated: those who have been caught not reporting correctly, the non-compliant, and, at the other side, those who can substantiate that their claims are correct and therefore get cleared (compliant). Thus, the reactions we see in the treatment group after the audit is made up by reactions in two different subsets. Simply measuring the average effect, as seen so far, would therefore conflate informative evidence about differences between the two groups, as also discussed by Gemmell and Ratto (2012) and Beer et al. (2015).

Table 5 shows descriptive statistics for the audited, when distinguishing between compliant and non-compliant taxpayers.

Table 5. Descriptive statistics, compliant, non-compliant and non-audited

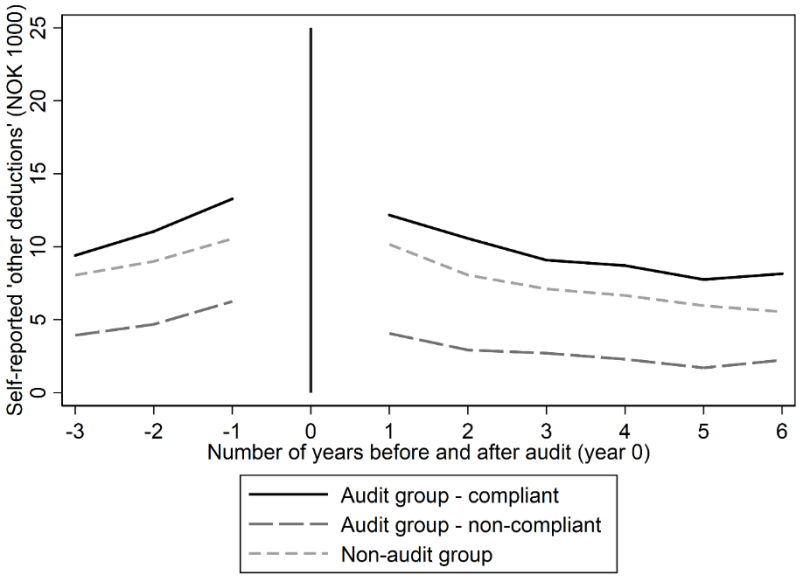
	Compliant	Non-compliant	Non-audited
Claimed "Other deductions"	22,275 (8,914)	23,153 (10,463)	23,104 (11,290)
Direct correction in "Other deduction" due to audit	-	25,948 (.)	-
Self-employed	0.25 (0.293)	0.14 (0.243)	0.21 (0.406)
Female	0.25 (0.458)	0.34 (0.490)	0.29 (0.454)
Age	52 (12.5)	43 (12.9)	48 (13.0)
Temporary work migrant	0.034 (0.199)	0.073 (0.243)	0.060 (0.237)
Total deductions	232,364 (79,232)	191,417 (70,882)	217,410 (205,798)
Third-party rep. deductions	153,390 (52,507)	128,443 (46,071)	141,596 (108,897)
Total gross income	859,141 (355,130)	617,731 (243,154)	770,830 (736,564)
Total third-party rep. gross income	655,057 (296,625)	526,447 (235,434)	607,691 (484,428)
Observations	2,238	1,238	26,775

Note: Standard deviations in parentheses

Figure 3 confirms that the post-audit behavior of the two groups is clearly different. Whereas taxpayers in the non-compliant group move considerably below the non-audited, the compliant taxpayers show the opposite behavior: after they have been audited and cleared, they report higher deductions than the non-audited. But Figure 3 also clearly depicts that this pattern is relatively parallel to the deduction behavior before the audit. Figure 3 demonstrates that the non-compliant taxpayers move into the group exposed to audit from a lower average deduction level, compared to the

compliant, and this pattern is repeated after the audit. The main question is (of course) if the audit itself could have affected the reactions.

Figure 3. Deductions before and after audit. Compliant and non-compliant taxpayers compared to the non-treated



Notes: Deductions of the year of the audit not reported in the figure

However, the patterns of Figure 3 also give some concern. At least they suggest that the taxpayers in the two groups initially are systematically different, which may represent a challenge for the identification of effects. The problem boils down to a concern about the non-audited representing a valid the counterfactual for both groups. Figure 3 indicates that there are systematic differences between the taxpayers in the two groups before the audit. Moreover, there may be unobserved factors that influence the selection into subgroups; they may for example differ with respect to aversion to risk or "inclination to criminal behavior". In this perspective Ratto and Gemmell (2012), Beer et al. (2015) and Advani, Elming and Shaw (2018) raise the question to what extent one can use the whole group of non-audited as defining the counterfactual behavior when obtaining results for the compliant group and the non-compliant group, respectively. Instead one would like to see post-audit deduction behavior for the compliant and non-compliant when the behavior in the two groups is compared to compliant and non-compliant types among the non-audited, respectively. But, of course, these two latent groups are not easily identified, as allocation primarily will be based on unobservables. As a response to this, Advani, Elming and Shaw (2018) suggest estimating effects based on the treated only, as they let the different groups identified by the audit be compared to themselves prior to the audit. Thus, they leave out non-audited in this of their analysis.

In the following we further discuss why using the "mixed" control group will lead to inconsistent estimates of the type-specific audit effects. To do this, we first introduce some necessary notations. Let $Q_i = 1$ denote that individual i is of type non-compliant and $Q_i = 0$ if the person is the

compliant type. As above, we have $D_i = 1$ if the individual audited and $D_i = 0$ if not. Let δ_1 and δ_0 be the DID estimates for the non-compliant and compliant taxpayers, respectively, using all the non-audited as the control group. Denote $\Delta Y_i(Q_i, D_i)$ as the difference in the deduction of individual i before and after the auditing. Then the DID estimator for the compliant group can be written as:

$$\begin{aligned}\delta_1 &= E[\Delta Y_i(Q_i = 1, D_i = 1) - \Delta Y_i(D_i = 0)] \\ &= E[\Delta Y_i(Q_i = 1, D_i = 1)] - E[\Delta Y_i(D_i = 0)] \\ &= E[\Delta Y_i(Q_i = 1, D_i = 1)] - pE[\Delta Y_i(Q_i = 1, D_i = 0)] - (1-p)E[\Delta Y_i(Q_i = 0, D_i = 0)],\end{aligned} \tag{5}$$

where p is the probability for individual i being a non-compliant taxpayer. Given the random assignment of audit, it can be consistently estimated. It is easy to see that δ_1 will be a consistent estimator of the type specific audit effect, $\gamma_1 = E[\Delta Y_i(Q_i = 1, D_i = 1) - \Delta Y_i(Q_i = 1, D_i = 0)]$, if and only if the following condition holds: $E[\Delta Y_i(Q_i = 1, D_i = 0)] = E[\Delta Y_i(Q_i = 0, D_i = 0)]$. That is, the change in outcome variable in absence of the treatment does not depend on the unobserved types. Or in other words, the common trend assumption holds.

The reason for us, at least as a start, to let the non-audited represent the counterfactual for both groups is that the common trend assumption seems to be fulfilled for both, see Figure 2.[15] Following Autor (2003), we have checked this more formally by regressing deductions in the two groups prior to the audit against time dummies and dummies for type of taxpayer, compliant or non-compliant taxpayer, denoted by Q_i (as established after the audit),

$$y_{it} = \alpha_i + \beta_1 Q_i + \sum_t \beta_{2t}\lambda_t + \sum_t \beta_{3t}\lambda_t Q_i + \varepsilon_{it} \quad t \in \{-3, -2, -1\}. \tag{6}$$

Recall that λ_t symbolizes the year-specific effects.

As we obtain clearly non-significant estimates of β_3 for all pre-audit years[16], we conclude that we see no signs of the common trend assumption being rejected. In the next subsection we shall show estimates for the compliant and the non-compliant taxpayers separately, letting the non-audited represent the counterfactual for both.

5.2. Effect of audit on subgroups

It follows from the discussion above that we employ the same specification as seen in Equation (3) to estimate average effects when the treatment group is divided into two subgroups. The specification is simply an extension of Equation (3) as we introduce a further distinction in the post-treatment years, dependent on the taxpayers been found guilty or not guilty. Thus, we let subscript j denote that the treated belong to the subgroups 1 and 2, compliant and non-compliant taxpayers,

[15] Although, it must be admitted that the pre-reform period is short.
[16] In practice this is done by running regressions with alternate year as the omitted.

(7) $$y_{it} = \alpha + \delta_{sj}\left(D_{it} \times k_{sj}\right) + \lambda_t + \mu_i + \varepsilon_{it} \quad s \in \{1,2,...,6\} \wedge j \in \{1,2\}\ .$$

It follows that estimates of effects of audits, δ_{sj}, are identical to the average effects except that we measure effects for compliant and non-compliant subgroups separately.

The results described in Table 6 suggest that there are large differences between taxpayers being told to adjust their claims and those experiencing that they have been cleared. Whereas the compliant taxpayers do not alter their deduction behavior after the audit, the non-compliant reduce their deductions substantially. Thus, this gives support for audits having a deterrence effect on the non-compliant, although, as noted above (see Section 2), these taxpayers have not been fined because of their unverified claims. Interestingly, the estimates of Table 6 give some support to the deterrence effect diminishing over time, and after six years the effect is no longer significant.

As noted in Section 3, we are not confident that all taxpayers belonging to the compliant-group have been aware of the audit. To the extent that this fact represents a substantial contribution to bias, these taxpayers will likely behave as belonging to the non-audited, and therefore potentially weaken the observed effects. With respect to the results for the compliant in Table 6, we see that point estimates shift between positive and negative values, which means that it is not clear to which direction a potential bias contributes.

Table 6. Effects of audit on post-audit deduction behavior. Compliant and non-compliant taxpayers

	Year after audit	Coefficient	Estimate	*t*-value
Compliant	First	δ_{11}	-400(611)	-0.65
	Second	δ_{21}	123(602)	0.21
	Third	δ_{31}	-384(603)	-0.64
	Fourth	δ_{41}	-302(617)	-0.49
	Fifth	δ_{51}	-85(638)	-0.13
	Sixth	δ_{61}	-90(794)	-0.11
Non-compliant	First	δ_{12}	-2,876***(589)	-4.88
	Second	δ_{22}	-1,858***(589)	-3.15
	Third	δ_{32}	-1,091*(622)	-1.75
	Fourth	δ_{42}	-1,045*(602)	-1.74
	Fifth	δ_{52}	-1,219**(611)	-1.99
	Sixth	δ_{62}	-405(740)	-0.55

Notes: Fixed effect estimation based on panel data 2008–2015. Robust standard errors in parentheses. $*p < 0.1, **p < 0.05, ***p < 0.01$

5.3. Spillover effects on other items

As noted in Section 4.3 an account of the costs and benefits of an audit should control for audits influencing the reporting on other items. In our case we may ask if the attention received in terms of

the check on the item "Other deductions" may cause the agents to adjust their subsequent filing behavior in general.

In order to some extent explore this issue further, we have estimated Equation (7) when replacing the dependent variable by gross income. Thus, if the dynamic audit effect spreads, as signified by results of Table 5, we expect to see similar patterns for gross income as well. However, we find no indications of spread to the gross income reporting.[17]

One may object that the third-party reporting schedule of Norway (see Section 3) prevent us from observing any behavioral effects along this channel, simply because most taxpayers do not control their income reporting. Given that the self-employed have wider scope for such behavioral adjustments we have estimated Equation (7) (with gross income as the dependent variable) only for the self-employed part. However, we neither obtain any statistically significant estimates when conditioning to the self-employed. But the point estimates are in accordance with what we would expect: the non-compliant taxpayers report higher income in the first three years after audit, whereas the point estimates suggest the opposite reaction among the compliant, as they reduce gross income after audit.

6. Robustness tests

6.1. Robustness test using partial identification method

Our results so far are derived from a DID framework critically dependent on the common trend assumption. In the following, we discuss results from empirical approaches that seek to obtain results under less restrictive conditions.

The first alternative is a version of partial identification method, where the ambition is to derive bounds to the group specific audit effects. In the case when the common trend fails, we could use (5) and some additional assumptions to derive the bounds for the true audit effects for compliants (γ_0) and non-compliants (γ_1), which are defined as

$$\gamma_0 = E[\Delta Y_i(Q_i = 0, D_i = 1)] - E[\Delta Y_i(Q_i = 0, D_i = 0)],$$

$$\gamma_1 = E[\Delta Y_i(Q_i = 1, D_i = 1)] - E[\Delta Y_i(Q_i = 1, D_i = 0)].$$

We can rewrite (5) as

$$\delta_1 = \gamma_1 + (1-p)(E[\Delta Y_i(Q_i = 1, D_i = 0)] - E[\Delta Y_i(Q_i = 0, D_i = 0)]). \quad (8)$$

Can we say anything on the sign of $E[\Delta Y_i(Q_i = 1, D_i = 0)] - E[\Delta Y_i(Q_i = 0, D_i = 0)]$? If we assume that it is non-positive; namely that in absence of audit the non-compliant taxpayers will not reduce their

[17] Results are not reported here but are available upon request.

deductions more than the compliant taxpayers, i.e. that the expression to the right of γ_1 in Eq. (8) is non-positive, then we have $\gamma_1 \geq \delta_1$. The audit effect for compliant taxpayers can be seen as

(9) $\delta_0 = \gamma_0 + p(E[\Delta Y_i(T_i = 0, D_i = 0)] - E[\Delta Y_i(T_i = 1, D_i = 0)])$.

Use the same assumption that we used to bound (8), we have $\gamma_0 \leq \delta_0$. We can then use the identity

(10) $p\gamma_1 + (1-p)\gamma_0 = \text{ATT}$

where ATT is the average effect of audit on the audited group, which is estimated in section (4.3). Thus,

(11) $\gamma_0 = \dfrac{\text{ATT} - p\gamma_1}{1-p} \leq \dfrac{\text{ATT} - p\delta_1}{1-p}$,

since we know that $\gamma_1 \geq \delta_1$. Similarly,

(12) $\gamma_1 = \dfrac{\text{ATT} - (1-p)\gamma_0}{p} \geq \dfrac{\text{ATT} - (1-p)\delta_0}{p}$.

Under these assumptions it follows that we can bound the type-specific audit effects, seen as

(13) $\gamma_0 \leq \min(\delta_0, \dfrac{\text{ATT} - p\delta_1}{1-p})$ and $\gamma_1 \in [\delta_1, \dfrac{\text{ATT} - (1-p)\delta_0}{p}]$.

Unfortunately, we were not able to bound γ_0 from below without imposing further assumptions. Hence, based on estimates reported in Section 4.3 and Section 5.2, we can establish the bounds for the group specific audit effects.

The bounds are reported in Table 7. Note that the bounds for the non-compliant group is quite tight, which can be considered as additional evidence that our common trend assumption is valid. For the complaint group, we obtain only the upper bounds, which are not significantly different from zero. In other words, we do not see any empirical evidence of positive audit effect for the compliants.

Table 7. Bounds for the effects of auditing

	Years after audit	Coefficient	Bound
Compliant	First	δ_{11}	≤ -400
	Second	δ_{21}	≤ 123
	Third	δ_{31}	≤ -384
	Fourth	δ_{41}	≤ -302
	Fifth	δ_{51}	≤ -85
	Sixth	δ_{61}	≤ -90
Non-compliant	First	δ_{12}	[-2,876, -2,776]
	Second	δ_{22}	[-1,858, -1,770]
	Third	δ_{32}	[-1,091, -1,043]
	Fourth	δ_{42}	[-1,045, -997]
	Fifth	δ_{52}	[-1,219, -1,158]
	Sixth	δ_{62}	[-405, -360]

6.2. Robustness test using a matching method

Another way to deal with the problem of latent types in the control group is to use some types of matching methods based on observed individual characteristics. By doing this we can reduce the potential bias generated by the fact that the control groups comprised of both compliants and non-complaints. However, this method will never fully solve this problem unless one can perfectly predict the latent type of a given individual. Nevertheless, we consider this as a useful check for the robustness of our main conclusion.

There are many different matching methods. Here we apply the Coarsened Exact Matching algorithm (CEM), see Iacus, King and Porro (2011) and use pre-audit control variables to obtain better balance between the treated and the control groups. Note that around 10% of audited individuals were not matched to anyone in the control group so they were excluded from the matched regression analysis.

Table 8 presents the results using only the matched sample. Compared with the non-matched sample, the estimated effects audits for the non-compliant groups are more clearly identified and the effects are larger. The point estimates for the compliant groups now are all positive but none of them are significant, except for the last year of period.

Table 8. Effects of audit on post-audit deduction behavior. Compliant and non-compliant taxpayers, matched sample.

	Year after audit	Coefficient	Estimate	*t*-value
Compliant	First	δ_{11}	513(671)	0.76
	Second	δ_{21}	1072(668)	1.61
	Third	δ_{31}	377 (664)	0.57
	Fourth	δ_{41}	891(673)	1.32
	Fifth	δ_{51}	1131(700)	1.61
	Sixth	δ_{61}	1662*(850)	1.95
Non-compliant	First	δ_{12}	-4,313***(599)	-7.20
	Second	δ_{22}	-3,161***(553)	-5.71
	Third	δ_{32}	-2,878*(606)	-4.75
	Fourth	δ_{42}	-2,877*(573)	-5.01
	Fifth	δ_{52}	-3,451**(606)	-5.70
	Sixth	δ_{62}	-2,537(763)	-3.32

Notes: Fixed effect estimation based on panel data 2008–2015. Robust standard errors in parentheses. Matching of sample carried out by Coarsened Exact Matching (CEM) algorithm. $*p < 0.1, **p < 0.05, *** p < 0.01$

7. Conclusion

It is crucial for the tax administration to know how much tax revenue that is generated by audits. Such calculations should not only account for the corrections made in the year of the control but must take into account that taxpayers being exposed to audits most likely adjust their behavior in the years after the audit. We find evidence that supports this conjecture. This result is in accordance with a target effect, simply stating that agents perceive that an audit is likely to be followed by another in the future. Thus, this suggests that there exists a deterrence effect of auditing.

Moreover, we benefit from being able to further explore the dynamics of tax audit by following the behavior of the compliant and non-compliant taxpayers. Some recent studies (Gemmell and Ratto, 2012; Beer et al., 2015) find that the compliant taxpayers show opposite reactions after an audit, suggesting that an approval through an audit may give counteracting effects. This can be explained by previously compliant taxpayers now inferring that the risk of future checks is low and thereby exploit the information provided by the tax administration to decrease the subsequent tax burden. However, we see no signs of such behavior in our analysis: the compliant taxpayers show the same deduction behavior as before the audit.

The distinction into compliant and non-compliant taxpayers makes the (likely) deterrence effect on the non-compliant to be more clearly identified, and it lasts for five years. Only in the last year of the observation period it has disappeared.

References

Advani, A., W. Elming, and J. Shaw (2017). The dynamic effects of tax audits. IFS Working Paper W17/24, Institute for Fiscal Studies.

Allingham, M. and A. Sandmo (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3-4), 323–338.

Alm, J. (2019). What motivates tax compliance. *Journal of Economic Surveys*, forthcoming.

Alm, J., B. R. Jackson, and M. McKee (2009). Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics*, 93, 392–402.

Andreoni, J., B. Erard, and J. Feinstein (1998). Tax compliance. *Journal of Economic Literature*, 36(2), 818–860.

Autor, D. H (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics*, 21, 1–42.

Beer, S., M. Kasper, E. Kirchler, and B. Erard (2015). Audit Impact Study, 2015 Annual Report to Congress, TAS Research and Related Studies.

Brinch, C., E. Hernæs, and Z. Jia (2017) Salience and Social Security Benefits. *Journal of Labor Economics*, 35(1), 265–297.

Blackwell, M., S. Iacus, G. King, and G. Porro (2009). Coarsened exact matching in Stata. *Stata Journal* 9(4), 524–546.

Bobek, D. D. and R. C. Hatfield (2003). An investigation of the theory of planned behavior and the role of moral obligation in tax compliance. *Behavioral Research in Accounting*, 15(1), 13–38.

DeBacker, J., B. T. Heim, A. Tran, and A. Yuskavage (2015). The impact of legal enforcement: An analysis of corporate tax aggressiveness after an audit. *Journal of Law and Economics*, 58(2), 291–324.

DeBacker, J., B. T. Heim, A. Tran, and A. Yuskavage (2018). Once bitten, twice shy? The impact of IRS Audits on Filer Behavior. Mimeo, University of South Carolina.

Dubin, J. A. (2007). Criminal investigation enforcement activities and taxpayer noncompliance. *Public Finance Review*, 35(4), 500–29.

Dubin, J. A., M.J. Graetz, and L. L. Wilde (1990). The effect of audit rates on the federal individual income tax, 1977–1986. *National Tax Journal*, 43(4), 395–109.

Dwenger, N., H. Kleven, I. Rasul, and J. Rincke (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *American Economic Journal: Economic Policy*, 8(3), 203–232.

Erard, B., and J. Feinstein (1994). The role of moral sentiments and audits perceptions in tax compliance. *Public Finance/Finances Publiques*, 49(Supplement), 70–89.

Feinstein, J. S. (1991). An econometric analysis of income tax evasion and its detection. *The RAND Journal of Economics*, 22(1), 14–35.

Fortin B., G. Lacroix, and M.-C. Villeval (2007). Tax evasion and social interactions. *Journal of Public Economics*, 91, 2089–2112.

Gangl, K., B. Torgler, E. Kirchler, and E. Hofmann (2014). Effects of supervision on tax compliance: Evidence from a field experiment in Austria. *Economics Letters*, 123, Issue 3, 378–382.

Gemmell, N. and M. Ratto (2012). Behavioral responses to taxpayer audits: Evidence from random taxpayer enquiries, National Tax Journal, 65(1), 33–57.

Hashimzade, N., G. D. Myles, and B. Tran-Nam (2013). Applications of behavioural economics to tax evasion. *Journal of Economic Surveys*, 27(5), 941–977.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46(4), 931–959.

Hernæs, E. and Z. Jia (2013). Earnings distribution and labour supply after a retirement earnings test reform, *Oxford Bulletin of Economics and Statistics*, 75(3), 410–434.

Iacus, S. M., G. King, and G. Porro (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106(493), 345–361.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651–92.

Luttmer, E. F. P. and M. Singhal (2014). Tax morale. *Journal of Economic Perspectives*, 28(4), 149–168.

Maciejovsky, B., E. Kirchler, and H. Schwarzenberger (2007). Misperception of chance and loss repair: On the dynamics of tax compliance. *Journal of Economic Psychology*, 28(6), 678–91.

Mendoza, J. P., J. L. Wielhouwer, and E. Kirchler (2017). The backfiring effect of auditing on tax compliance. *Journal of Economic Psychology*, 62, 284–294.

Mittone, L. (2006). Dynamic behaviour in tax evasion: An experimental approach. *The Journal of Socio-Economics*, 35(5), 813–835.

Mittone, L., F. Panebianco, and A. Santoro (2017). The bomb-crater effect of tax audits: Beyond the misperception of chance. *Journal of Economic Psychology*, 61(C), 225–243.

Myles, G. D. and R. A. Naylor (1996). A model of tax evasion with group conformity and social customs. *European Journal of Political Economy*, 12(1), 49–66.

Ratto, M., R. Thomas, and D. Ulph (2013). The indirect effects of auditing taxpayers. *Public Finance Review*, 41(3), 317–333.

Sandmo, A. (2005). The theory of tax evasion: A retrospective view. *National Tax Journal*, 58(4), 643–63.

Slemrod, J., M. Blumenthal, and C. Christian (2001). Taxpayer response to an increased probability of an audit: Evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, 79(3), 455–483.

Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives*, 21(1), 25–48.

Slemrod, J. (2018). Tax compliance and enforcement. NBER working Paper 24799, National Bureau of Economic Research.

Snow, A. and R. S. Warren, Jr. (2005). Tax evasion under random audits with uncertain detection. *Economics Letters*, 88(1), 97–100.

Snow, A. and R. S. Warren, Jr. (2007). Audits uncertainty, Bayesian updating, and tax evasion. *Public Finance Review*, 35(5), 555–571.

Tversky, A. and D. Kahneman (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.

Yitzhaki, S. (1974) A note on: "Income tax evasion: a theoretical analysis". *Journal of Public Economics*, 3, 201–202.