Statistisk sentralbyrå
Statistics Norway

Do audits improve future tax compliance in the absence of penalties? Evidence from random Audits in Norway

TALL SOM FORTELLER

Shafik Hebous, Zhiyang Jia, Knut Løyland, Thor O. Thoresen, and Arnstein Øvrum

DISCUSSION PAPERS

943

Discussion Papers No. 943, November 2020
Statistics Norway, Research Department

Shafik Hebous, Zhiyang Jia, Knut Løyland,
Thor O. Thoresen, and Arnstein Øvrum

Do audits improve future tax compliance in the absence of penalties? Evidence from random Audits in Norway

Abstract:
The Norwegian Tax Administration operated multi-year random audits of personal income tax returns. We exploit this exceptional randomized setup to estimate the effects of tax audits on future compliance explicitly distinguishing between dynamic responses of compliant and noncompliant audited taxpayers. A priori, the literature has suggested two competing effects: A post-audit deterrence effect—whereby audits prompt taxpayers to comply in subsequent years—or a "bomb-crater" effect—whereby audits lower taxpayers' subjective probability of detecting future evasion and hence weaken compliance. Our results show improved future compliance for five post-audit years by those that were found noncompliant in the audits, despite the absence of penalty, suggesting that it is not the monetary payment per se that carries a deterrence effect. Those that were found compliant, however, show no signs of behavioral adjustments. Although the findings are consistent with the deterrence effect, mainly stemming from being caught of wrongdoing rather than a penalty, we argue that there is also a "learning" effect with the important implication that better information for taxpayers critically complements tax audits.

Keywords: Tax administration, tax evasion, tax compliance, tax audits, administrative data

JEL classification: H26, C23

Acknowledgements: This research has been carried out as a collaboration between the Norwegian Tax Administration and Oslo Fiscal Studies (Department of Economics, University of Oslo). Support from the Norwegian Research Council to Oslo Fiscal Studies is gratefully acknowledged. We thank Edwin Leuwen, Monique de Haan, participants in the "Modelling tax policy and compliance workshop" (TARC, University of Exeter), Skatteforum 2019, Annual Congress of the International Institute of Public Finance (Glasgow) for helpful comments. The views expressed here are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

Address: Shafik Hebous, International Monetary Fund. Email: Shebous@imf.org
Zhiyang Jia, Statistics Norway, Research Department. E-mail: zhiyang.jia@ssb.no
Knut Løyland, Norwegian Tax Administration. Email: knut.loyland@skatteetaten.no
Thor O. Thoresen, Statistics Norway and Oslo Fiscal Studies.
Email: thor.olav.thoresen@ssb.no
Arnstein Øvrum, Norwegian Tax Administration. Email: arnstein.ovrum@skatteetaten.no

Discussion Papers comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no/en/forskning/discussion-papers
http://ideas.repec.org/s/ssb/dispap.html

ISSN 1892-753X (electronic)

Sammendrag

På hvilken måte påvirker kontroller utført av skattemyndighetene atferden til de skattepliktige i årene etter at de har blitt kontrollert? I årene 2009, 2010 og 2011 utførte skatteetaten kontroll av skattepliktige som har brukt feltet «Andre fradrag» på skattemeldingen. Nærmere bestemt har de plukket ut skattepliktige som har ført «Andre fradrag» i intervallet fra 5 000 til 50 000 kroner. Denne fradragsposten kan anvendes til utgifter som ikke kan plasseres under andre fradrag. Dette kan for eksempel være utgifter til hjemmekontor, diverse avgifter i forbindelse med kapitalinvesteringer og donasjoner til særskilte formål.

Kontrollen kan føre til at den skattepliktige får endret skattemeldingen av Skatteetaten i det gjeldende år (myndighetsfastsetting). Vi er ikke opptatt av denne direkte effekten av kontrollen, men hvordan kontrollen påvirker personens atferd i de påfølgende inntektsårene (indirekte effekter). Er det slik at folk rapporterer mindre eller mer utgifter (på feltet «Andre fradrag») i årene etter å ha blitt kontrollert? Vi måler indirekte effekter i opptil 6 år etter kontroll. Det kan vi gjøre fordi skattemyndighetene kontrollerte et tilfeldig utvalg på omtrent 10 prosent av dem som oppfylte vilkårene for (fradrag mellom 5 000 og 50 000 kroner). Vi kan dermed se på atferden til de tilfeldig utvalgte kontrollerte og sammenlikne dem med de andre skattepliktige som oppfylte vilkårene for kontroll.

Det betyr også at vi får to grupper av personer som har blitt kontrollert: Dem som har blitt kontrollert og klarert og dem som har fått myndighetsfastsetting. Skattemeldingen til personer i den siste gruppen blir korrigert, men overtredelsen utløser ikke nødvendigvis noen tilleggsskatt. Det siste kan motiveres ved at overtredelsene er små og en kan ikke utelukke at kravene som ikke har blitt godtatt skyldes misforståelser av regelverket.

Det er flere teorier for hvordan folk opptrer etter å ha blitt kontrollert. Først og fremst kan en kontroll ha en slags «avskrekkingseffekt», som betyr at folk rapporterer mer korrekt etter at skattemyndighetene har vist interesse for selvangivelsen deres fordi de antar at det øker sannsynligheten for å bli undersøkt neste gang også. Men det er også utviklet en teori som beskrives som «bombekrater-effekten», som viser til at soldater i krig gjemmer seg i et bombekrater under antakelsen av at en bombe ikke slår ned på eksakt samme punkt i løpet av kort tid. Det betyr at personer i den siste gruppen vil kunne øke sine fradrag etter å ha blitt tatt i uregelmessigheter, fordi de antar at sannsynligheten for ny kontroll er liten når de allerede har blitt kontrollert.

Våre resultater tyder på at det er avskrekkingseffekten som gjør seg gjeldende her. Vi finner at den kontrollerte gruppen som helhet reduserer bruken av feltet «Andre fradrag» i det første året etter kontrollen. Punktestimatene tyder på at effektene er negative (og avtakende) også i de påfølgende år, men estimatene er ikke signifikant forskjellige fra null. Vi ser klarere effekter når vi deler de kontrollerte i de to undergruppene som fikk og ikke fikk rettet sin likning. I den første gruppen ser vi ingen effekt av kontrollen, mens vi finner signifikante negative effekter i den siste gruppen i fem år etter kontrollen (effekten i det sjette året er også negativ, men ikke signifikant forskjellig fra null).

Våre resultater er viktige for å måle hva en oppnår med å kontrollere skattytere. De viser at en kontroll ikke kun innebærer at skatteprovenyet øker som følge av at selvangivelsen korrigeres i det året kontrollen er utført, men at de skattepliktiges atferd også endres i de påfølgende årene. Dette er viktig informasjon on betydningen av kontrollvirksomheten i skatteetaten.

Vår tolkning av resultatene er dessuten at avviste påstander om fradrag fra de skattepliktige like gjerne kan skyldes misforståelser av regelverket som at skattyteren med overlegg prøver å oppnå urettmessig lavere skatt. Hovedgrunnen til en slik tolkning er de relativt store atferdsendringene vi observerer til tross for at en relativt liten andel av de skattepliktige har fått utskrevet tilleggsskatt.

1 Introduction

A central premise of tax audits is enhancing tax compliance. This insight—dating to the seminal work by Allingham and Sandmo (1972)—has been predominantly viewed in a static manner: a higher probability of detecting tax evasion, and thus paying a penalty, lowers the expected utility from concealing income, ceteris paribus.[1] The static nature of this prediction and the direct effects of audits on revenues in the year of the audit (through adjusting income and penalties) have been extensively studied and empirically confirmed.[2] However, with a few exceptions, there is little evidence, thus far, on the dynamic effects of tax audits—i.e., to what extent do tax audits encourage or discourage compliance in post-audit years? This question is crucial for an optimal tax administration as it is one element in the evaluation of the cost and effectiveness of administrative interventions (Slemrod and Keen, 2017).

This study estimates the effects of tax audits on future tax compliance using novel Norwegian administrative personal tax records data for about 30,000 individuals and random audits in the period 2009–2011. Our database contains rich information on all taxpayers eligible for the audit—not only those that were selected for the audits—and all are observed for up to six years after the audit and up to three years before the audit. Our empirical analysis has two important features. First, the research design is based on random audits. Second, it contributes to the literature by studying how audits can impact future compliance following misreporting even in the absence of penalties. The analysis of "real-life" random tax audits using administrative tax data provides relevant estimates for tax authorities and contributes to supporting one of two competing broad theoretical predictions.

Theoretically, the post-audit compliance behavior is ambiguous as it ultimately depends on taxpayers' perception of the detection probability, which can go in either direction. Taxpayers may perceive that an audit in this year is likely to be followed by other audits (and thus higher detection probability) in subsequent periods—a deterrence effect (e.g., Hashimzade, Myles and Tran-Nam, 2013)—, but it is also equally foreseeable that an audit today prompts agents to perceive a significant decline in the probability of being audited again in the aftermath of this audit—the "bomb-crater" effect[3] (e.g., Maciejovsky, Kirchler and Schwarzenberger (2007) and Mittone, Panebianco and Santoro, 2017).

[1] There is a range of further determinants of tax compliance beyond the Allingham-Sandmo model, which fall under the broad umbrella of tax moral (Luttmer and Singhal, 2014). See also Lederman (2018) for a discussion of voluntary compliance.

[2] For surveys, see, e.g., Andreoni, Erard, and Feinstein (1998), Alm (2019), and Slemrod (2019).

[3] The notion that a "bomb" is unlikely to strike exactly at the same place again.

Either way, this classical reasoning assumes that there is a monetary penalty. However, if, ex post, misreporting was caught and not subjected to penalty, then theory becomes even more elusive offering no guidance on how the absence of penalties interacts with forming priors about future detection probabilities, which again can go in either direction. Thus, empirical evidence based on administrative data is necessary to shed light on the evolution of the effects of audits on taxpayers' behavior over time.

Norwegian taxpayers receive fully prepopulated income tax returns that they digitally file by a mere mouse click by simply approving the tax return. This largely automatic and digitalized process relies on an extensive and highly developed third-party reporting system. Taxpayers do have the option of entering missing information and making some amendments. One particular item of interest in the tax form is "Other deductions" for claiming deductions that are not already recorded. This particular item has been subject to random audits by the tax administration, which generated the dataset of this study.

Our empirical strategy exploits the random nature of tax audits by the Norwegian Tax Administration, which, critically, randomizes both the assignment into a treated group (audited) and a control (non-audited). Usually, audits are based on risk scores. Hence, the challenge that typically faces this type of analysis—and most studies on audits—is the nonrandom selection for tax audits in practice leading to a severe selection bias. In this study, given the random assignment, our main identification strategy for obtaining consistent and unbiased estimates is a difference-in-difference (DiD) research design. The validity of our approach is based on the common trend assumption, which we can reasonably motivate by simple plots of self-reported deductions by the three groups: unaudited, audited-compliant, and audited-noncompliant taxpayers.

Our DiD main results suggest that audited taxpayers, overall, reduce their claimed income deductions in the post-audit years. However, this is an average effect for two groups as audit outcomes split audited taxpayers into compliant and noncompliant. Estimates for the compliant and the noncompliant taxpayers separately indicate that the compliance effect is driven mainly by the noncompliant group, lasting for five years but it decreases over time. The decrease in reported deductions is 12 percent in the immediate post-audit year and 5 percent in the fifth post-audit year, measured against the average claimed deduction in the year of the audit. Note that adjusted incomes (after the audits) were subjected to the tax but in principle without imposing penalties. In contrast, the compliant group shows no statistically significant change in reporting during post-audit years. One issue is that the unaudited group potentially contains both compliant and noncompliant taxpayers. We openly discuss this aspect and show that the common trend assumption remains the key aspect for valid group-specific estimates. Additionally, using different methodologies, we obtain similar results to those from the DiD design. In particular, we provide estimates from: i) a less restrictive approach that derives bounds' estimates of the group-specific audit effect following the methodology in Manski and Pepper (2018); and ii) matching techniques.

Thus, based on our results we can rule out the bomb-crater effect of the audits. Yet, it is a striking finding that noncompliant taxpayers did comply following the audit even without penalizing them. This suggests that monetary penalty per se does not play a significant deterrence effect but rather being notified of wrongdoing is what carries a significant dynamic compliance effect. We motivate that this result is consistent with a "learning effect"—i.e., the dynamic compliance effect is, at least in part, prompted by taxpayers' learning as a result of audits.

In practice, misreporting in the personal income tax return—especially for this type of items—may not be entirely driven by elaborate tax evasion scheme but can also reflect uncertainty about the tax treatment or inadvertent errors.[4] While it is not possible to explicitly distinguish deliberate evasion from acting under uncertainty, strictly speaking unverified claims still comprise tax evasion.[5] However, evasion (or misreporting) due to uncertainty largely reflects a "wishful thinking"—following the notion that "maybe I can deduct it without substantiating it or maybe not, but hopefully I get away with it".[6]

From a tax collection point of view, the estimated dynamic effect is what ultimately matters irrespective of the motivation to comply in post-audit years—whether it is purely "deterrence" or "learning". And our main contribution is providing evidence supporting the deterrence effect of the audits enhancing compliance. However, our results also have the important implication that it is not always the penalty what matters but improving taxpayers' information and strengthening the communication of tax policy are important complements to deterrence strategies, thereby comprising important aspects of optimal tax administration.

[4] Especially given the modest—albeit non-negligible—amounts involved in this study. As explained in Section 2, the maximum amount that can be spared by tax evasion by using "Other deductions" in this study is 1,650 USD.

[5] It is an active decision to report deductions in the tax return without confirming their eligibility or being able to substantiate these deductions.

[6] See, e.g., Heller and Winter (2020) and Mayraz (2011).

Our study contributes to various strands of the literature. First, a few studies empirically examine the reaction of taxpayers to audits. Slemrod et al. (2001) find that taxpayers report higher income after being warned about future audits of their income returns. DeBacker et al. (2015) study corporate behavior in the US and find that firms reduce tax payments immediately after audits but increase payments gradually afterwards. DeBacker et al. (2019) find that the effect of audits on future tax payments in the US is short-lived without third-party information. In contrast, Advani et al. (2019) find that third party information does not predict whether a taxpayer is compliant in the UK. Kleven et al. (2011) find in a field experiment in Denmark that tax evasion is close to zero for income subject to third-party reporting, in stark contrast to self-reported income. Gemmell and Ratto (2012) distinguish between noncompliant and compliant taxpayers in the UK and report evidence suggesting that the "past experience" rather than the threat of audits enhance future tax compliance by the noncompliant. Beer et al. (2020) study the dynamic compliance effect for self-employed taxpayers in the US, but the selection of audits is based on a risk score. They find that compliance depends on the audit outcome and that non-audited taxpayers reduce their reported income in post-audits years. We contribute to this literature by i) using high-quality tax return data with actual random audits; ii) looking at the role of taxpayer learning in enhancing compliance beyond the threat of penalty; and iii) distinguishing between the responses of compliant and noncompliant taxpayers, which with the exception of Beer at al. (2020) and Gemmell and Ratto (2012) has not been explicitly studied.

Second, several studies have examined compliance behavior in a laboratory environment. For example, Alm, Jackson, and McKee (2009) find that pre-announced audit rates improve compliance. This finding is broadly in line with theorical predictions in Snow and Warren (2007), suggesting that as unaudited taxpayers update expectations about the probability of future audits (i.e., Bayesian learning) tax evasion increases. Our empirical paper complements this literature with evidence from high quality administrative data.

Third, as surveyed by Luttmer and Singhal (2014), a strand of the literature extends the Allingham-Sandmo model in several directions accounting for compliance factors beyond enforcement, among other things, moral sentiments of guilt and shame (Erard and Feinstein, 1994) and social conformity effects (Myles and Naylor, 1996; Fortin, Lacoix, and Villeval, 2007). Somewhat relatedly, Beck and Jung (1989) theoretically model uncertainty about tax lability and predict that, as a result, the impact on compliance is ambiguous. Despite a general recognition that uncertainty and errors are important factors that explain noncompliance (e.g., Slemrod, 2007), there have been no empirical studies, thus far, looking at tax audits in this context. Alm, Jackson, and McKee (1992) in a laboratory experiment find that uncertainty tends to increase tax compliance, but the effect in principle can go in either direction. In this context, we contribute to the literature by providing suggestive evidence that learning by taxpayers (and thus lower uncertainty) improves tax compliance, feeding into the broad intrinsic motivation of compliance.

The remainder of this paper is organized as follows. Section 2 describes the Norwegian institutional setting that produced the audit data. Section 3 presents descriptive statistics distinguishing between compliant and noncompliant taxpayers. Section 4 presents estimates for the overall behavioral effect of tax audits, for compliant and noncompliant taxpayers, and includes results of several robustness checks. Section 5 concludes.

2 Institutional Setting

Norway adopts a tax system that is close to a dual income tax, thereby broadly business and capital incomes (such as from interest) are subject to a flat tax rate (which was 28 percent in 2011 and is 22 percent as of 2020) whereas gross employment income is subject to a progressive tax schedule with a top personal income tax (PIT) rate of 47.8 percent in 2011 (46.4 percent in 2020). The top PIT rate roughly kicks in at a level of income equal to a multiple of 1.6 of the average wage. Given that income after deduction is subject to the flat rate at 28 percent, the value of a deduction in terms of reduced tax burden is 0.28 times the deduction.

The Norwegian third-party information system—based on data from employers, the financial sector, and others—is advanced, almost fully eliminating the need for contact between taxpayers and the tax administration. Taxpayers simply receive a digitally fully prepopulated tax return for approval. Prefilled tax returns can be amended by taxpayers. For example, charitable donations are deductible up to a threshold in the taxation of ordinary income. The recipients of donations (i.e., recognized charitable organizations) report the individual donations directly to the tax authorities. Of course, in case errors occur or incomes or deductions are not reported, taxpayers can make amendments to the income tax return directly through the internet without the involvement of administrative staff from the tax authorities.

The analysis of this paper focuses on a particular item in the income tax return called “Other deductions”, which is frequently used to report additional deductions not already recorded through the third-party reporting system. Common categories of deductions under this item are fees related to capital income—including management fee, and stock exchange account—, expenses for home office, and charitable donations to Civil Society Organizations or scientific entities.[7] This item is filled in by wage earners and the self-employed.

The nature of this self-declared item requires substantiating the claimed deductions. Correspondingly, from a tax administrative standpoint, audits are required as taxpayers can make substantiated or unsubstantiated (illegal) claims. All taxpayers with a claim under "Other deductions" above 50,000 (8,300 USD)[8] were audited. Among taxpayers who have claimed "Other deductions" in the range from 5,000 to 50,000 NOK (approximately 570–5,700 USD), a subsample of approximately 10 percent is randomly assigned for further audits. The random assignment and audit rules were used by the Norwegian Tax Administration for only three years: 2009, 2010 and 2011, and thus this study uses information from auditing in these years.

Noncompliant taxpayers were informed that their tax returns have not been approved by the tax authorities after assessment and that errors were found in the deduction item. Some compliant taxpayers were informed about the inspection by the tax administration while others may go through the process without notification, if they already have provided all the necessary documentation needed. However, many become aware of the audit because they have been asked to provide additional information. For example, if a deduction of 50,000 NOK was denied by the tax authorities, a 28-percent flat rate applies, implying an extra tax of 14,000 NOK (about 1,650 USD). Misreporting in this item does not exclusively imply deliberate evasion but can also reflect inadvertent errors or uncertainty about the tax treatment. In principle, there is a penalty on top of the regular tax (given by the flat rate) in the case of a deliberate tax evasion. Note that unverified deductions constitute, in principle, tax evasion (even under "uncertainty"), but in this study, *ex post*, given the relatively modest amounts involved, no penalty was imposed.[9]

---

[7] Other, less frequently used categories include judicial costs, newspaper subscriptions, moving costs, and dental costs. Some of these deductions will be limited to one year, while others (annual fees) will be repeated for several years.

[8] Exchange rates for 2010.

[9] Only about 1.5 percent of the noncompliant taxpayers were subjected to a monetary penalty. In the Law, the penalty rate was 20 percent of the benefit obtained from evasion. For example, the evaded amount is 50,000 NOK—at a tax rate of 22 percent, the benefit from evasion is 11,000 NOK. In case a penalty if imposed, it would be 2,200 NOK.

3 Audit Results

3.1 Data Description

We base our analysis on three waves of audits on “Other deductions”, in 2009, 2010, and 2011. For the taxpayers who are eligible for the audits, we have information on the particular deduction claimed and a set of other individual characteristics including gender, age, total gross income, third part reported gross income, total deductions (including the item we study), third party reported deductions, and self-employment status.[10] The observation period starts in 2008 and lasts until 2015. That is, we observe taxpayers in 1–3 years before the audit and in 3–6 years after the audit.

The auditing process generates two distinctively different groups among the treated: those who have been caught not reporting correctly, the noncompliant, and those who can substantiate that their claims are correct and therefore get cleared (compliant). Around 36 percent of the taxpayers are found to have misreported their deductions.

Table 1. Descriptive Statistics for Compliant, Noncompliant, and Non-audited (Averages), 2008–2015

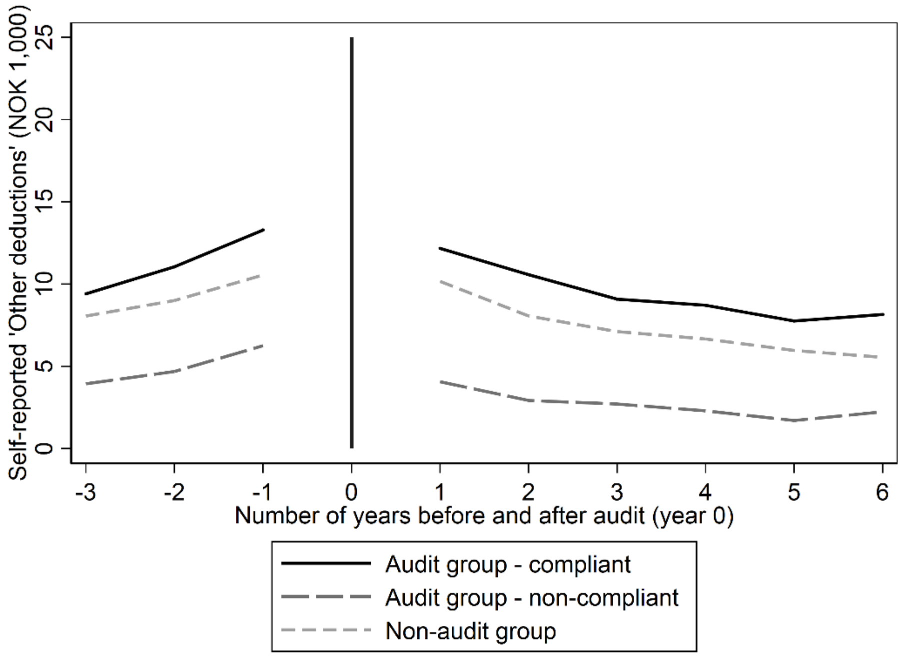
	Compliant	Noncompliant	Audited in total	Non-audited
Claimed “Other deductions” (in the year of the audit)	22,275 (8,914)	23,153 (10,463)	23,091 (11,520)	23,104 (11,290)
Direct correction due to audit	–	25,948 (22,182)	–	–
Self-employed†	25 (29.3)	14 (24.3)	21 (41.0)	21 (40.6)
Female†	25 (45.8)	34 (049.0)	28 (45.0)	29 (45.4)
Age (years)	52 (12.5)	43 (12.9)	49 (13.0)	48 (13.0)
Temporary work migrant†	3.4 (19.9)	7.3 (24.3)	4.8 (21.0)	6.00 (23.7)
Total deductions	232,364 (79,232)	191,417 (70,882)	217,780 (214,060)	217,410 (205,798)
Third-party rep. deductions	153,390 (52,507)	128,443 (46,071)	144,505 (116,134)	141,596 (108,897)
Total gross income	859,141 (355,130)	617,731 (243,154)	773,161 (733,130)	770,830 (736,564)
Total third-party rep. gross income	655,057 (296,625)	526,447 (235,434)	609,252 (480,575)	607,691 (484,428)
Observations	2,238	1,238	3476	26,775

Notes: Standard deviations are in parentheses. † Figures are in percent. The rest of the values are in NOK except for age (years).

[10] An individual is classified as a self-employed if the income from the business is at least 10 000 NOK, and even if the individual is a wage earner as well.

Table 1 shows descriptive statistics for the non-audited and the audited distinguishing for the latter group between compliant and noncompliant taxpayers. The average claimed ‘other deductions’ is close to 23,000 NOK (3,800 USD) for the three groups. Appendix A presents the distribution of deductions by groups of taxpayers. The average correction of deductions for the noncompliant group is about 26,000 NOK (4,300 USD). Additionally, the summary statistics reveal some notable differences between the two subgroups of the audited. For example, the compliant taxpayers are older and richer than individuals in the noncompliant group. Moreover, in the noncompliant group there are more self-employed (25 percent vs. 14 percent) and somewhat surprisingly more females (34 percent vs. 25 percent) than in the compliant group. This is in contrast to some studies that find that males are more likely to evade taxes than females (Torgler and Valev, 2010). The self-employed are typically the focus of tax evasion studies because they have more scope to misreport income than wage earners that are subject to wage withholding taxes. However, this item of deductions is self-reported for both the self-employed and the non-self-employed, and thus the self-employed do not have more flexibility to misreport this item.

Figure 1. Average Deductions before and after the Audits: Compliant, Noncompliant and Non-audited



Notes: Year of the audit is denoted by 0 on the horizontal axis. Diagram is representative for the observations used in the regressions, i.e., year 0 is excluded. Based on observations 1–3 years before the audit and in 3–6 years after the audit.

3.2. Deduction Behavior by Group

Figure 1 shows the trend in the deduction behavior for all subgroups: non-audited, audited-compliant, and audited-noncompliant. Recall that the three waves of audits were in 2009–2011 (coded as year 0). We observe taxpayers for up to six years after the audit and three years pre-audit. Note that Figure 1 shows the average values for the pre- and post-t_0, whereas Table 1 displays the average values of “Other deductions” for yeart_0.

Figure 1 shows that individuals in the noncompliant group (on average) have much lower claimed deduction level than individuals in the compliant group, despite that the levels are almost the same at the year of audit (as shown in Table 1). In Appendix C, we rationalize different levels of deductions as observed in Figure 1. In the absence of audits, compliance behavior is less correlated over time, but non-compliant taxpayers claim potentially illegitimate deductions only if the amount is relatively high. Thus, in terms of level, it can be well the case that unverified claims are lower than legitimate deductions.

One important question is: Which characteristics can explain the assignment into compliant or noncompliant group? We estimate a simple Probit model of the form:

$$Pr(\text{ non-compliance}) = \Phi(\beta_i x_i), \tag{1}$$

where x_i represents the observed characteristics and Φ is the standard normal Cumulative Distribution Function (CDF). While the results from such a model cannot be interpreted as causal, they are informative, at least regarding two aspects. First, this kind of information can be used to help to design a more efficient audit program by targeting those who are more likely to be noncompliant. Second, these results may shed lights on possible mechanism underlying the compliance behavior. Table 2 presents the average marginal effects of a range of characteristics on the probability of being noncompliant. The main findings broadly confirm the summary statistics in Table 1. For example, being self-employed and male have significant negative effects on the likelihood of being a noncompliant. Also, the estimation results suggest that immigrant status does not have a significant effect on the probability of belonging to the noncompliant group (after controlling for other variables).

Table 2. The Probability of Being Noncompliant: Average Marginal Effects

	AME	Standard Error
Self-employed	-0.07	0.021
Female	0.04	0.017
Age	-0.01	0.001
Temporary work migrant	-0.02	0.349
Total deductions	0.04	0.052
Third-party rep. deductions	-0.26	0.101
Total gross income	-0.04	0.016
Total third-party rep. gross income	-0.04	0.023
Observations		3,476
McFadden R^2		0.107

Note: Results from Probit model estimation. Total deductions, third-party reported deductions, gross income and third-party reported gross income are measured in million NOK.

4 Estimation of the Dynamic Audit Effects

4.1 Hypothesis Development

Generally, reported income (y) in year t is a function of perceived probability of detection (p), penalty (ϕ), and a vector of other variables (Z): $y = f(p, \phi, Z)$. The Allingham-Sandmo model unambiguously predicts that a higher penalty rate or a higher perceived probability of detection lowers tax evasion, thereby increasing reported income—i.e., $\partial y/\partial \phi > 0$ and $\partial y/\partial p > 0$. In a dynamic setup, p itself can be a function of past values of variables including whether an audit was conducted in previous years: $p_t=f(Audit_{t-1}, \ldots)$. The effect of past audits on p_t crucially affects future compliance in post-audit years but it is ultimately an empirical question as theory offers ambiguous predictions. The marginal effect of the last year audit on reported income is given by:

$$\frac{\partial y_t}{\partial Audit_{t-1}} = \frac{\partial f}{\partial p_t}\frac{\partial p_t}{\partial Audit_{t-1}} + \frac{\partial f}{\partial \phi_t}\frac{\partial \phi_t}{\partial Audit_{t-1}}$$

The second term in our study is zero—i.e., $\frac{\partial \phi_t}{\partial Audit_{t-1}} = 0$, because fines are not used. The deterrence hypothesis predicts that $\frac{\partial p_t}{\partial Audit_{t-1}} > 0$ (and thus $\partial y_t/\partial Audit_{t-1} > 0$) whereas the bomb-crater hypothesis predicts that $\frac{\partial p_t}{\partial Audit_{t-1}} < 0$ (and thus $\partial y_t/\partial Audit_{t-1} < 0$).

The first contribution of our analysis is to estimate the marginal effects $\partial y_t/\partial Audit_{t-1}$, $\partial y_{t+1}/\partial Audit_{t-1}$, $\partial y_{t+2}/\partial Audit_{t-1}$, …, for up to six post-audit years, thereby testing which effect dominates.[11] The second

[11] Since we study deductions, we look at the negative of (a fraction of) y.

contribution of our empirical analysis stems from the fact that misreporting was (ex post) not penalized enabling us to test whether audited taxpayers improve their compliance even if ϕ_{t-1} is zero. This is a novel aspect and relevant for at least two reasons. First, it generally indicates whether the monetary penalty per se plays a significant deterrence effect or if it is rather being notified of wrongdoing that carries a significant deterrence effect (i.e., $\frac{\partial y_t}{\partial Audit_{t-1}} > 0$ despite that $\partial y / \partial \phi = 0$). Whether a higher monetary penalty implies a higher compliance, as the Allingham-Sandmo model predicts, is a contested empirical question. While in the static Allingham-Sandmo model, without expected penalty, evasion becomes pervasive (abstracting from tax morale motives), in a repeated interaction setup without penalty, taxpayers still have to form priors about the future probability of detecting evasion which may or may not entail penalty. In sum, it is unclear whether taxpayers internalize the no-penalty outcome of the audit in their future compliance decision necessitating empirical analysis.

Second, if future compliance improves after detecting misreporting despite the absence of a penalty, then the bomb-crater effect can be unambiguously rejected. This would be clear-cut because if the perceived detection probability has declined with no (or lower perceived) penalty the result would be unambiguously higher evasion. Yet, if we can rule out the bomb-crater effect hypothesis, the improved compliance effect would be a mixture of a strict deterrence effect and a learning effect that eliminates uncertainty. To see this, consider the knowledge of the tax system as one variable in Z, call it *W*. The marginal effect of past audit on reported income becomes:

$$\frac{\partial y_t}{\partial Audit_{t-1}} = \frac{\partial f}{\partial p_t} \frac{\partial p_t}{\partial Audit_{t-1}} + \frac{\partial f}{\partial W_t} \frac{\partial W_t}{\partial Audit_{t-1}}$$

The above equation shows that reported income can increase following an audit even without a change in the perceived probability p (i.e., when $\frac{\partial f}{\partial p_t} \frac{\partial p_t}{\partial Audit_{t-1}} = 0$) through the learning term $\frac{\partial f}{\partial W_t} \frac{\partial W_t}{\partial Audit_{t-1}}$. This learning term captures the case when the audit prompts taxpayers to know about the legitimacy of their deductions eliminating any uncertainty. Without the audit, wishful-thinking and acting under uncertainty would be present as long as there are taxpayers with incomplete information about the validity of their deductions. While it is challenging—if not impossible—to empirically explicitly distinguish between the contribution of both terms to misreporting, it is the general nature of the deduction item under consideration together with the no-penalty that motivate a role of uncertainty and taxpayers' learning in our study.

4.2. Overall Audit Effect

The overall, average, post-audit effect can be estimated either by comparing the outcome of the audited and unaudited—given that the assignment of audit is random—or using a simple DiD design with unaudited taxpayers as the control group. These two methods are in principal equivalent, but the latter has the advantage that it can be applied when we want to separately estimate effects for the compliant and noncompliant groups (as discussed below).

We estimate the overall post-audit effect of the audit in year zero on the amount of deductions of individual i using the following DiD specification:

(2) $$Deduction_{is} = \alpha_i + \lambda_t + \theta D_i + \sum_{k=1}^{6} \delta_k \left[D_i \times \mathbb{1}_{(s=k)}\right] + \varepsilon_{ist} \qquad \forall\, s \in \{-3, -2, -1, 1, \dots, 6\},$$

where s measures the distance in years to the year of the audit. The binary regressor, D_i, takes the value 1 if the taxpayer was audited in year $s = 0$; and it is zero otherwise. α_i denotes a set of individual fixed effects that captures unobserved heterogeneous characteristics of taxpayers and λ_t represents the calendar year effect that captures all factors that affect all taxpayers in a year. The year of the audit is excluded from the dataset used in the estimation. Thus, δ_s measures the treatment effect of audit at a specific year s after the audit.

Table 3 reports the estimated audit effects on deductions in six post-audit years. There is statistically significant negative effect in the first year after the audit suggesting that “Other deductions” decline by approximately 1,300 NOK, on average. For the rest of the post-audit years (i.e., for $s > 1$), point estimates suggest that taxpayers reduce their reporting of deductions because of the audit and depict a diminishing effect overtime, but they are insignificant. As the results indicate that there is a negative shift in the mean deduction after the audit, in Appendix B we examine whether the shape of the deduction distribution has been changed. The results indicate that the audit affects deduction claims on both the intensive and extensive margin (i.e., the number of individuals claiming “Other deductions).

Table 3. Effects of Tax 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.

$$* \, p < 0.1, ** \, p < 0.05, *** \, p < 0.01$$

4.3. Composition of the Non-Audited

The above estimated average effect is informative, but it masks heterogenous responses of those who were found compliant and noncompliant as a result of the audit, calling for separate estimates for both groups. The identification of a group-specific effect is, however, challenging because the behavior of taxpayers in the non-audited group is latent —i.e., it is not necessarily exclusively representing the compliant group or the noncompliant group. It is rather a “mixed” control group, consisting of individuals who would be identified as both compliant and noncompliant taxpayers had they been audited.

We discuss here the question as to what extent it is possible to use the whole group of non-audited as the control group for obtaining treatment estimates for the compliant and the noncompliant. Advani, Elming, and Shaw (2019) use a before-after setup comparing the compliance behavior prior to and after the audit. This method overcomes a possible endogeneity problem, but it only allows identifying the effect for the compliant group (and not for the noncompliant group). Moreover, it does not have a control group for the post-audit period even for the compliant group.[12]

We show, below, that the common trend (for the non-audited and for the group-specific trends) suffices to obtain consistent reliable estimates. Let $Q_i = 1$ denote that individual i is of type noncompliant and $Q_i = 0$ denote the compliant type. As above, we have $D_i = 1$ if the individual is audited and $D_i = 0$ if not. Let δ_1 and δ_0 be the DiD estimates for the noncompliant and compliant taxpayers, respectively, using all the non-audited as the control group. $\Delta Y_i(Q_i, D_i)$ denotes the difference between the post-audit and the pre-audit deduction of individual i. Then the DiD estimator for the noncompliant group can be written as:

[12] Gemmell and Ratto (2012) simply used the non-audited as the control group for both the compliant and noncompliant groups.

$$\begin{aligned}(3)\qquad \delta_1 &= E[\Delta Y_i(Q_i = 1, D_i = 1) - \Delta Y_i(D_i = 0)] \\ &= E[\Delta Y_i(Q_i = 1, D_i = 1)] - E[\Delta Y_i(D_i = 0)] \\ &= E[\Delta Y_i(Q_i = 1, D_i = 1)] - qE[\Delta Y_i(Q_i = 1, D_i = 0)] - (1-q)E[\Delta Y_i(Q_i = 0, D_i = 0)],\end{aligned}$$

where q is the probability for individual i being a noncompliant taxpayer. Given the random assignment of audit, it can be consistently estimated. 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. In our data, the common trend assumption seems to hold for both groups (Figure 1), supporting using the non-audited group as a control group for both types of the audited taxpayers (compliant or noncompliant).

Following Autor (2003), we formally test the common trend assumption by regressing deductions in the two groups prior to the audit, $s \in \{-1, -2, -3\}$, against time dummies and dummies for type of taxpayer, compliant or noncompliant taxpayer, denoted by Q_i (as established after the audit), as follows:

$$(4)\qquad Deduction_{i,s} = \alpha_i + \lambda_t + \beta Q_i + \sum_{k=-3}^{-1} \xi_k \left[Q_i \times \mathbb{1}_{(s=k)}\right] + u_{ist} \qquad \forall\, s < 0,$$

where α_i is a set of individual fixed effects and λ_t represents year fixed effects. We estimate Equation (4) omitting years in an alternate manner, $s \in \{-1, -2, -3\}$. We obtain statistically insignificant estimates of ξ_k for all pre-audit years (results are not reported). Hence, we conclude that there is no statistical support for the hypothesis of rejecting the common trend.

Thus, overall, given the above discussions and results, we argue that the non-audited group is a valid control group. Nonetheless, we return to this methodological challenge below, in terms of results from a less restrictive partial identification method.

4.4 Separate Estimates for the Compliant and Noncompliant

Following the above discussion, we extend specification (2) to estimate group-specific effects. In particular, we introduce a further distinction in the post-treatment years distinguishing between two types of taxpayers: those who were able to substantiate the claimed deductions versus those who were not. We estimate Equation 2 separately for compliant and noncompliant taxpayers. In this setup, the δ_k give the group-specific estimates of the effects of audits on reported deductions in each post-audit year k (separately for the compliant and for noncompliant groups).

Table 4 displays the estimation results. The findings suggest that there are large differences between cleared taxpayers and those that were requested to adjust their claims due to the lack of sufficient substantiation. First, compliant taxpayers do not alter their deduction behavior after the audit. Estimated coefficients are statistically insignificant, and they change signs. Second, noncompliant taxpayers reduce their deductions by 2,876 NOK (480 USD) in the first year and 8,089 NOK (1,340 USD) over the five post-audit years. The compliance effect diminishes over time and turns insignificant in the sixth post-audit year. Thus, overall, the results are not consistent with a bomb-crater effect as they cannot be reconciled with a lower subjective probability, but they rather lend support to an improved compliance effect for the noncompliant.

The results are striking as noncompliant taxpayers were not fined because of their unverified claims. This suggests that audits can have positive dynamic effects on compliance even in the absence of penalties. Broadly, this means that it is not the monetary penalty as such that preserves the deterrence effect of audits—i.e., $\frac{\partial f}{\partial p_t} \frac{\partial p_t}{\partial Audit_{t-1}}$ can be larger than zero even without a penalty. Moreover, the result is consistent with a learning effect, $\frac{\partial f}{\partial W_t} \frac{\partial W_t}{\partial Audit_{t-1}} > 0$, suggesting that misreporting is at least in part prompted by "wishful thinking" or more generally by uncertainty about the tax treatment of some deductions. Thus, the decline in claiming unsubstantiated deductions by taxpayers can be reconciled with a higher perceived probability of audits and with an unchanged perceived probability of audits via a learning effect.

Table 4. Effects of Audit on Post-audit Deduction Behavior

	Year after audit	Coefficient	Estimate	t-value
Compliant	First	δ_{01}	-400(611)	-0.65
	Second	δ_{02}	123(602)	0.21
	Third	δ_{03}	-384(603)	-0.64
	Fourth	δ_{04}	-302(617)	-0.49
	Fifth	δ_{05}	-85(638)	-0.13
	Sixth	δ_{06}	-90(794)	-0.11
Noncompliant	First	δ_{11}	-2876***(589)	-4.88
	Second	δ_{12}	-1858***(589)	-3.15
	Third	δ_{13}	-1091* (622)	-1.75
	Fourth	δ_{14}	-1045* (602)	-1.74
	Fifth	δ_{15}	-1219** (612)	-1.99
	Sixth	δ_{16}	-405 (740)	-0.55

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

4.5 Robustness Tests

Establishing Bounds Based on A 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 an empirical approach that seek to obtain results under less restrictive conditions, a version of a partial identification method, where the ambition is to derive bounds to the group-specific audit effects. The same underlying idea is used by Manski and Pepper (2018) under the name the Bounded Variation Assumptions approach.

When the common trend fails, we can use Equation (3) and some additional assumptions to derive bounds for the true audit effects for the compliant group (γ_0) and for the noncompliant group (γ_1). The true effects can be defined as

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

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

Further, we introduce two assumptions, with respect to average behavior, that our empirical approach relies on. Firstly, in absence of audit the noncompliant taxpayers will not reduce their deduction claims more than the compliant taxpayers, and, secondly, there are more reductions for those who get caught than for the compliant taxpayers not being audited.

In the following we formalize how bounds can be derived based on these relatively mild assumptions. The exercise is primarily helpful in order to clarify in which direction one would expect results to move when not relying on a subgroup common trend. Then, one should be aware that the first assumption, that the noncompliant taxpayers will not reduce their deduction claims more than the compliant taxpayers, basically states that the γ_1 is not above the average treatment effect of the noncompliant, δ_{1s}in Table 4.

Nevertheless, let us see how the bounds can be derived. The two assumptions imply that we have $E[\Delta Y_i(Q_i = 1, D_i = 0)] \geq E[\Delta Y_i(Q_i = 0, D_i = 0)]$ and

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

This means that we have the following conditions for audit effect of the noncompliant group:

(7) $$\gamma_1 = \delta_1 - (1-q)(E[\Delta Y_i(Q_i = 1, D_i = 0)] - E[\Delta Y_i(Q_i = 0, D_i = 0)]) \leq \delta_1,$$

which gives

(8) $$\gamma_1 = \frac{1}{p}\{\delta_1 + (1-q)(E[\Delta Y_i(Q_i = 1, D_i = 1)] - E[\Delta Y_i(Q_i = 0, D_i = 0)])\} \geq \frac{\delta_1}{q}.$$

Thus, we bound the true treatment effect for the noncompliant group as

$\frac{\delta_1}{q} \leq \gamma_1 \leq \delta_1$.

For the compliant group we correspondingly have,

$\delta_0 = \gamma_0 + q(E[\Delta Y_i(Q_i = 0, D_i = 0)] - E[\Delta Y_i(Q_i = 1, D_i = 0)])$.

When we use the same assumption as employed to restrict γ_1 in Eq. (8), we get $\gamma_0 \geq \delta_0$. To obtain the upper bound, we can then use the identity

$q\gamma_1 + (1-q)\gamma_0 = \text{ATT}$,

where ATT is the average effect of audit on the audited group, of which estimation results already have been obtained. Thus,

$$\gamma_0 = \frac{\text{ATT} - p\gamma_1}{1-q} \leq \frac{\text{ATT} - \delta_1}{1-q},$$

which follows from $\gamma_1 \geq \frac{\delta_1}{q}$.

Under these assumptions the bounds the type-specific audit effects can be seen as

(9) $$\gamma_0 \in [\delta_0, \frac{\text{ATT}-\delta_1}{1-q}] \text{ and } \gamma_1 \in [\frac{\delta_1}{q},\ \delta_1].$$

Hence, based on estimates reported in Section 4.1 and Section 4.3 we obtain empirical estimates of the bounds for the group specific audit effects. It follows from our two conditions that the point estimates, δ_{0s} and δ_{1s} in Table 4, represent the lower and upper bound for the compliant and noncompliant, respectively. Intuitively, the tightness of the bounds for γ_1 is an increasing function of the share of individuals belonging to the noncompliant group (q). When there is are no noncompliant individuals in the population, that is when$q = 0$, there is no information in the data to identify γ_1, while the exact identification is obtained when $q = 1$. In this case the interval is reduced to a single point.

The bounds are reported in Figure 2 (without standard errors), showing that bounds are relatively wide for the noncompliant taxpayers. However, as one would expect, given the two assumptions that found the basis for obtaining them, the results point to possible directions if one leaves the common trend assumption. If anything, noncompliant taxpayers may reduce their deduction claims more after being audited, whereas the upper bounds of the compliant signify a possibility for approval.

Matching Method Results

To further check of the robustness of the results, we estimate a new model with a new control group obtained from a matching procedure.[13] In particular, we apply the Coarsened Exact Matching algorithm (CEM) and use pre-audit control variables to obtain better balance between the treated and the control groups. Approximately, 10 percent of the audited individuals were not matched to anyone in the control group, so they were excluded from the matched regression analysis.

[13] See Iacus, King, and Porro (2011).

Figure 2. Bounds for the Effects of Auditing

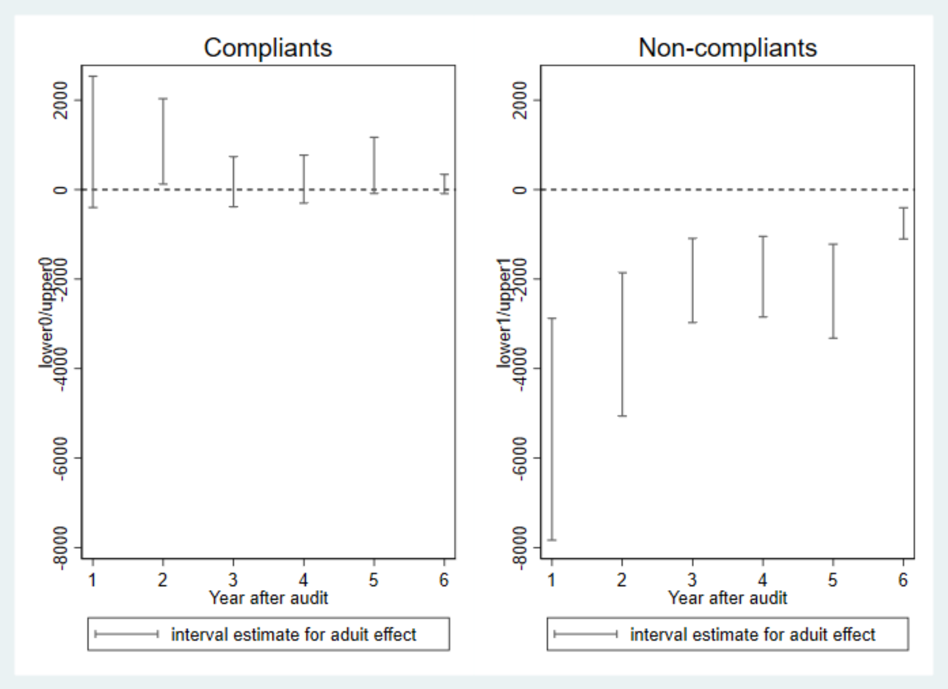


Table 5. Effects of Audit on Post-Audit Deduction Behavior: Matched Sample

	Year after audit	Coefficient	Estimate	*t*-value
Compliant	First	δ_{01}	514(672)	0.76
	Second	δ_{02}	1073(668)	1.61
	Third	δ_{03}	378(665)	0.57
	Fourth	δ_{04}	892(674)	1.32
	Fifth	δ_{05}	1131(701)	1.61
	Sixth	δ_{06}	1663*(850)	1.96
Noncompliant	First	δ_{11}	-4,314***(599)	-7.20
	Second	δ_{12}	-3,162***(554)	-5.71
	Third	δ_{13}	-2,878***(606)	-4.75
	Fourth	δ_{14}	-2,877***(574)	-5.01
	Fifth	δ_{15}	-3,452***(606)	-5.70
	Sixth	δ_{16}	-2,538***(764)	-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.$*\ p < 0.1$,$**\ p < 0.05$,$***\ p < 0.01$

Table 5 presents the results using only the matched sample. Compared with the non-matched sample, the estimated effects audits for the noncompliant 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. Moreover, we also obtained results for propensity score matching, which are very close to the results reported in Table 5.

Spillover Effects on Other Items

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 explore this issue further, we estimate Equation (2) separately for compliant and noncompliant taxpayers but using gross income as the dependent variable. If the dynamic audit effect spreads to the reporting of income to, we expect to see similar patterns as for "Other deductions" for gross income too. However, we find no indications of spread to the gross income reporting. Results are not reported here but are available upon request.

5 Conclusion

The effectiveness of audits is one crucial element of an efficient tax administration. In terms of assessing the revenue implication of audits, to draw the big picture, the calculation should not only account for tax adjustments made in the year of the audit, but also future tax adjustments triggered by behavioral responses to the initial audit. Based on data from random audits by the Norwegian Tax Administration, the findings of this study suggest that audited taxpayers reduced their claimed income deductions in the post-audit years, thereby raising their reported income and hence compliance.

Moreover, the analysis suggests that the increased future compliance effect is driven by the those that were audited and prompted to correct their tax returns. The decrease in their reported deductions is 12 percent in the first post-audit year, then it gradually decreases reaching 5 percent in the fifth post-audit year. However, no dynamic reaction was found for those that were audited and their tax returns were approved by the tax authorities without adjustments.

While this outcome, in general, can be explained by increasing taxpayers' subjective probabilities of future audits and detection (i.e., a deterrence effect), the analysis suggests that the dynamic improvement in compliance can be triggered even in the absence of penalties broadly in line with a learning effect. This implies that improving the information set of taxpayers is one of the key aspects of an efficient tax administration.

References

Advani, A., W. Elming, and J. Shaw (2019). The dynamic effects of tax audits. Working Paper No. 414, University of Warwick.

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*, 33(2), 353–388.

Alm, J., B. R. Jackson, and M. McKee (1992). Institutional uncertainty and taxpayer compliance. *The American Economic Review*, 82(4), 1018–1026.

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.

Beck, P.J. and W.-O. Jung (1989). An economic model of taxpayer compliance under uncertainty. *Journal of Accounting and Public Policy*, 8(1), 1–27.

Beer, S., M. Kasper, E. Kirchler, and B. Erard (2020). Do audits deter or provoke future tax noncompliance? Evidence on self-employed taxpayers. *CESifo Economic Studies*, 1–17.

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

Czibor, E., D. Jimenez-Gomez, and J. A. List (2019). The Dozen Things Experimental Economists Should Do (More of). *Southern Economic Journal*, 86(2), 371–432

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. *Journal of Law and Economics*, 61(1), 1–35.

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.

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

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.

Heller, Y, and E. Winter (2020). Biased-belief equilibrium, *American Economic Journal: Microeconomics* 12 (2), 1–40.

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.

Lederman, L. (2018). Does enforcement reduce voluntary tax compliance?, *Brigham Young University Law Review* 627.

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.

Manski, C. and J. V. Pepper (2018). How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions. *The Review of Economics and Statistics*, 100(2), 232–244.

Mayraz, G. (2011): Wishful thinking. CEP Discussion Paper No 1092, Centre for Economic Performance, London School of Economics, London.

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., 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.

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. and M. Keen (2017). Optimal tax administration. *Journal of Public Economics*, 152, 133–142.

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

Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, 57(4), 904–954.

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

Torgler, B. and N. T. Valev (2010). Gender and public attitudes toward corruption and tax evasion, *Contemporary Economic Policy* 28(4), 554–568.

Appendix A. Distribution of "Other Deductions"

Figure A1. Distribution of "Other deductions" among audited and non-audited. The year of the audit

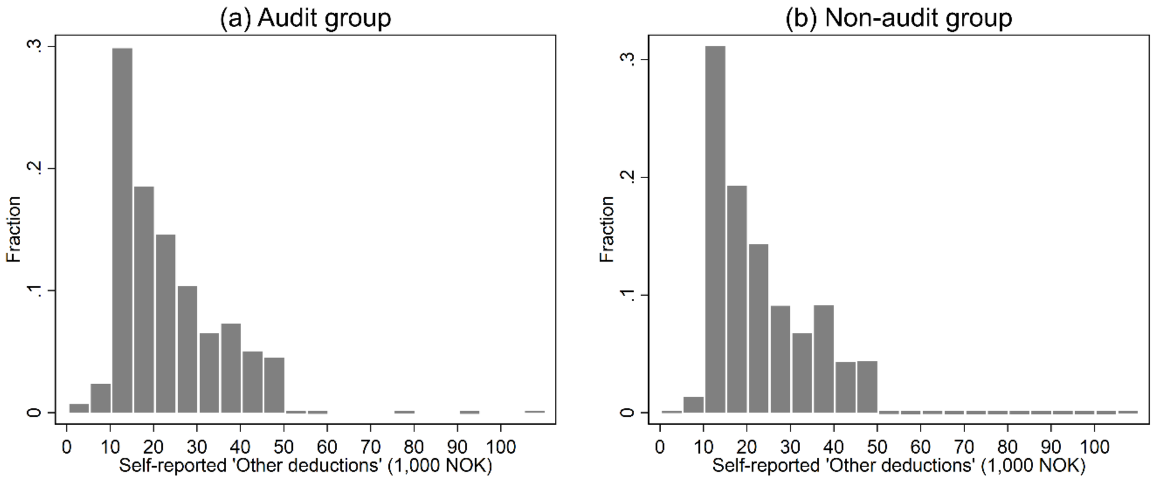
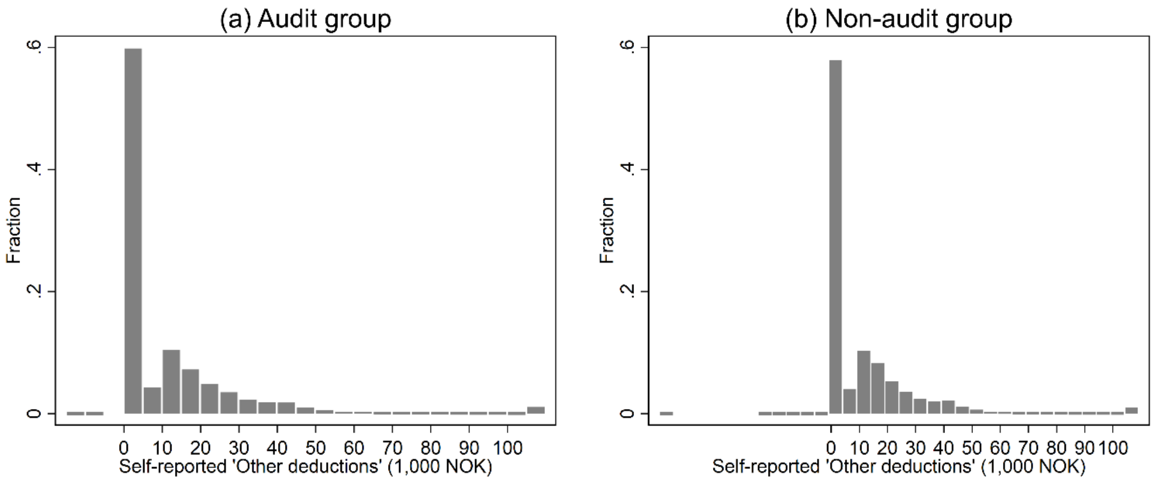


Figure A2. Distribution of "Other deductions" among audited and non-audited. Averages over all years used in the empirical analysis



Appendix B. Distribution of Treatment Effect

As the results indicate that there is a negative shift in the mean deduction after the audit, it is informative to examine to what extent the shape of the deduction distribution has been changed. Following Hernæs and Jia (2013) and Brinch, Hernæs and Jia (2017), we look at the Complementary Cumulative Distribution Functions (CCDF), $\bar{F}(y|X) = Pr(Y > y|X)$, before and after audit. In particular, 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 for individual i:

(B1) $$Pr(y_{it} > y_k) = F(\alpha_k + X_{it}\beta_k + \lambda_{tk} + \gamma_k D_i + \delta_{tk} D_i \times 1(t > 0)),$$

where X_{it} denotes individual characteristics and Frepresents the logit function. We estimate a series of logit models on the probability of claiming deduction above y_k, where it varies from NOK 0 to NOK 100,000 by increments of NOK 5,000. 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_k: $Pr(y_{it} > y_k|D_{it} = 1, X_{it}) - Pr(y_{it} > y_k|D_{it} = 0, X_{it})$.

Figure B1 is based on 21 separate estimations, one for each of the deduction levels, $y_k \in \{0, 5000, \dots, 95000, 100000\}$, shows the estimated marginal effects with 95 percent confidence envelops over these different deduction levels for the first year after the audit ($t = 1$). 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 the 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 B1.

Figure B1. Audit effects on the distribution of deductions, the year after audit

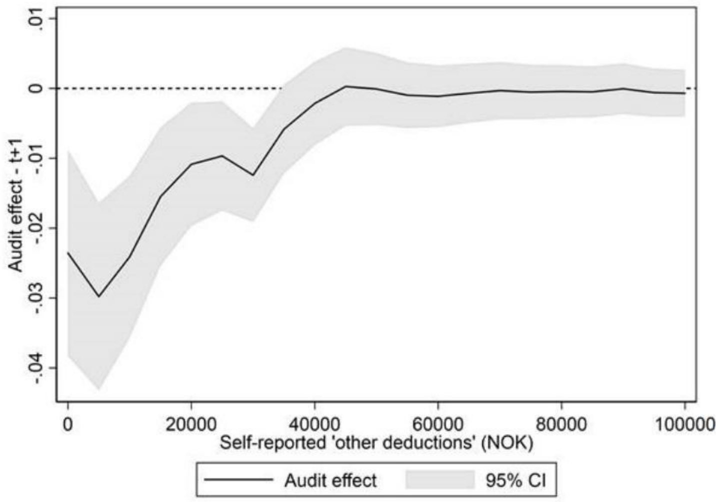


Table B1. Estimating the Deduction Interval after Audits (Results for the first year after audit)

	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

Appendix C: Tax Evasion under Uncertainty

Model Setup

A main message of the present study is that our results may not comply with intentional fraud behavior but could result some type of "learning". In the following we set up a simple model where we explain the observed differences in deduction claiming behavior before and after the audit observed in the compliant and noncompliant groups. In the model, individuals do not cheat intentionally. The behavior is instead a result of individuals mistakenly claim illegitimate deductions which they are unsure about themselves – for instance due to a lack of understanding of the tax rules.

For a given taxpayer, there is a set of potential tax deductions that they may claim. Among these deductions, some are "risk free", as the taxpayer knows for certainty that they are legitimate. Other deductions are "risky" in the sense that the taxpayer is unsure whether they are legitimate or not. The taxpayer chooses which deductions to claim.

Next, the tax authority conducts a random audit among taxpayers whose claimed deductions are above a given level, denoted as C. While taxpayers are aware that their claimed deductions may be audited, they do not know the rule of the audit selection. Some deductions are limited to one year while others are repeated for several years. Thus, we assume that the "risk-free" claim consists of a time invariant part, λ_i, and time varying part, ε_{it}, which we assume to be i.i.d. over time and individual with $E[\varepsilon_{it}] = 0$.

The "risky" claim $u_{it} > 0$ is independent over time and uncorrelated with ε_{it}. The independence assumption is restrictive but not essential for our main results. We assume that there is a subject belief probability $0 < p_i < 1$ that the risky claim is legitimate, which can be seen as a proxy of a self-evaluation of tax system knowledge.

The taxpayer will always claim "risk free" deductions. There is, however, a positive cost if the tax payer's claimed deduction is audited and found to be illegitimate. Thus, she will only claim the risky deduction if its amount is above a certain threshold. This threshold should depend on, among others, on two subjective probabilities: the probability of it being legitimate, p_i, and the probability of being audited, q_i. In our model, taxpayers may claim when they are not certain about the legitimacy of the claim, and it is argued that the decision is guided by the size of the loss (increased tax burden). Of course, such behavior can be a result of several misconceptions, such as "wishful thinking" bias (on wishful thinking bias see Mayraz, 2011)).

Model implications

For simplicity, we have assumed that all taxpayers are observably identical. In other words, the implication is valid when we control for observed characteristics. Let us consider a three-period model, $t \in [-1,0,1]$, where $t = 0$is the year of audit and $t = -1$ and $t = 1$ are the year before and after the audit, respectively. At any given year t, taxpayer i's claimed deduction is denoted as y_{it}. Among taxpayers, there are two types of individuals. Type I are those who claim only the risk-free deductions ($G_{it} = 0$), while Type II individuals claim both types ($G_{it} = 1$). Total claim can then be written as

$$y_{it} = \lambda_i + \varepsilon_{it} + G_{it} u_{it}$$

We understand immediately that the noncompliant group consists of only type II individuals who claimed "risky" deductions, whereas the compliant group consists of both types. Since u_{it} is uncorrelated over time, then group membership dummy G_{it} is uncorrelated over time. This implies that the compliance behavior is not correlated over time. Namely, being a noncompliant at a given audit gives no additional information on her compliant behavior in years prior to the audit. This result is quite strong, since it rules out intentional fraud and implies that the deduction behavior will be similar for these two groups prior to the audit. This is consistent with what we observe, i.e., that time trends for the deduction claimed prior to the audit are parallel. Note that this is not true for the behavior after audit, since the audit will change the key parameters governing the model, as we will return to below.

Implication 1. Noncompliant group has lower deductions pre-audit

We claimed that the deduction behavior prior to the audit would not be different for the compliant and noncompliant groups. However, we do observe that there are level differences even after we control for observed characteristics. In the following, we will show that this is due to the special eligibility criteria used in the audit we study.

Since we assume that the level of risk-free amount is uncorrelated with the risky amount, we see immediately that

$$E[y_{it} | G_{i0} = 0] < E[y_{it} | G_{i0} = 1].$$

For the audit we study, the taxpayer is eligible to audit only when the total deduction level is above a certain level. This implies that

$$E[\lambda_i|G_{i0} = 0, \text{qualified for audit}] = E[\lambda_i|G_{i0} = 0, y_{it} > C]$$
$$> E[\lambda_i|G_{i0} = 1, y_{it} > C] = E[\lambda_i|G_{i0} = 1, \text{qualified for audit}].$$

Thus, we have

$$E[\lambda_i|\text{compliance at } 0] > E[\lambda_i|\text{non-compliance at } 0].$$

Together with the assumption that u_{it} is uncorrelated over time, we have

$$E\left[y_{i,-1}|\text{compliance at } 0\right] > E\left[y_{i,-1}|\text{non-compliance at } 0\right].$$

The intuition is rather straightforward: suppose the there are two individuals who claim the same amount of deduction, one is complaint and the other is noncompliant. Since individuals only claim a "risky" deduction when the amount is high, the noncompliant will have lower time invariant risk-free claim than the compliant. Thus, this explains the pattern seen in Figure 1 prior to the audit.

Implication 2. Noncompliants adjust deductions downward post-audit

As mentioned above, there are mainly two key parameters which define the deduction claim behavior: the subjective belief on his own knowledge of the tax rules, proxied by belief probability, p, and the probability of getting audited, q.

After experiencing that their "risky" deductions have been corrected, the taxpayers would likely adjust downward their subject belief probability, p_i. On the other hand, the probability of getting audited, q_i, could go either way. If there is no "bomb crater" effect, or at least the reduction in p_i dominates a possible increase in q_i (bomb-crater), we will see reductions in the claimed deductions after the audit.

Implication 3: Compliants may adjust their deductions in either direction post-audit

For the compliant, the direction is less clear. Assume for now that they are aware of the fact they have been audited and all deductions are found legitimate. It is possible that they will adjust upward the subject belief, p_i. On the other hand, they may also adjust upward the audit probability, q_i. Thus, the overall effect could go either way.

Overall, the above-mentioned implications of the model are consistent with what we found in empirical analysis. While we cannot really test the basic assumptions of our model directly against data, the empirical results do show some inconsistencies with the theory that taxpayers evade when they have the chances. What we found points to another possible sources of tax noncompliance behavior, namely the complicated tax rules. Similar problems have been found in other cases where economic policies induce unintended outcomes, see for example Brinch, Hernaes and Jia (2017) for an example in the pension policy.