Comparing compliance effects and costs of hard and soft tax enforcement

Andreas Kotsadam, Knut Løyland, Oddbjørn Raaum, Gaute Torsvik
and Arnstein Øvrum

April 15, 2024

Abstract

In this study, we assess the effectiveness of two distinct tax enforcement strategies implemented among a subset of taxpayers characterized by high self-reported tax deductions. The selected taxpayers were randomly assigned to either undergo a correspondence audit or receive a letter encouraging a re-evaluation of their claimed deductions. Our findings reveal that the intervention letter and the audit resulted in a reduction in self-reported deductions, both in the intervention year and in the following year. An audit incurred approximately five times the cost of a letter and yielded a five times larger reduction in self-reported deductions. Hence, evaluated according to welfare costs per unit of tax income, measured by the marginal value of public funds (MVPF), both instruments perform roughly equal. We discuss to what extent MVPF serves as a sufficient statistic for informing practical decisions on whether to prioritize scaling up the letter or the audit in tax enforcement strategies.

JEL-codes: D04, H26, H83

Keywords: tax audits, tax revenue, tax compliance, income tax.

1 Introduction

Tax administrations employ various strategies to ensure accurate financial reporting and payment of owed taxes. These efforts generally fall into two categories: control and encouragement strategies. Control strategies involve monitoring and enforcement through tools such as audits, examinations, and investigations. Additionally, tax authorities use less intrusive and less expansive ex-ante enforcement policies, such as information campaigns, reminders, and encouragements, to promote voluntary compliance (OECD, 2019).

Alternative enforcement strategies typically have different costs as well as different effects on taxpayer compliance. Audits are more expensive to conduct than sending information notes to taxpayers with suspicious filings. On the other hand, audits may uncover a higher fraction of those with irregular filings, and we also know that audits have positive effects on subsequent tax compliance (Kleven et al., 2011; Advani et al., 2021; DeBacker et al., 2018; Hebous et al., 2023; Løyland et al., 2024). The evidence for the compliance effects of less expensive and softer enforcement policies is more mixed, even in the short run (see Alm, 2019; Slemrod, 2019 and Pomeranz and Vila-Belda, 2019). To our knowledge, there exists no empirical research that explicitly compares tax compliance effects and enforcement costs for both hard (control) and soft (encouragement) interventions. A credible comparison should be based on the same target group under identical circumstances.

This study attempts to close this knowledge gap. In many countries, the filing of tax information is increasingly done by third parties such as employers and financial institutions (Kleven et al., 2011). Taxable income, however, also depends on deductions that are often self-reported. Our field study was designed and implemented by the Norwegian Tax Administration (NTA). The study contrasts two interventions; a standard desk-based audit and an encouragement note. Random assignment provides credible estimates of the instantaneous and subsequent compliance effects which can be compared with their respective implementation costs.

An important feature of our study is that we consider alternative enforcement policies within the same population of personal taxpayers with suspicious tax returns, i.e., taxpayers that are typically targeted by operational audits. The NTA employed a prediction model to identify a target group that exhibits a high probability of making substantial unwarranted self-reported deductions. Among the 15 000 individuals with the highest risk scores, taxpayers were randomly assigned to one of three groups. The first group received a letter encouraging them to review their self-reported deductions (Letter), the second group had their deductions audited by a tax auditor (Audit), and the third group underwent

no special intervention (business as usual). Our research investigates the tax filing behavior of these groups in the year of the intervention and the subsequent year, with a specific focus on self-reported deductions in taxable income.

In addition to presenting findings on the cost-effectiveness of both hard and soft tax enforcement strategies within a policy-relevant group of high-risk filers, our study makes a separate contribution to the existing literature on soft tax enforcement. Previous research has demonstrated that making audits and detection probabilities more salient can influence reporting in the short term (Slemrod et al., 2001; Kleven et al., 2011; Fellner et al., 2013; Bott et al., 2020), as can information on public disclosure and prison sentences (**?**). However, studies grounded in social norms or tax morality have often found limited effects on tax compliance (Hallsworth, 2014). While some recent papers indicate increased compliance with such appeals (Hallsworth, 2014; Holz et al., 2023; Bott et al., 2020), others suggest possible backfiring (De Neve et al., 2021).

Using survey data from Uruguay, Bergolo et al. (2020) find that tax evasion is essentially uncorrelated with tax morale. In a meta-analysis of 40 studies, Antinyan and Asatryan (2019) find that nondeterrence nudges are ineffective, whereas deterrence nudges seem to have effects.[1] In a large field experiment from Belgium, De Neve et al. (2021) find that simplified tax filing information increased the number of individuals that filed their report on time and the effect persisted a year after the intervention.

We find that both the audit and the letter treatment lowered self-reported deductions in the year of the intervention. Both policies also reduced self-reported deductions in the subsequent tax year. The accumulated compliance effect over the two years is largest for the audit. Both the audit and the letter generate positive net tax revenue, but the audit approximately five times the net tax revenue generated by the letter. When evaluating these interventions within the framework of the marginal value of public funds (MVPF), both interventions yield public funds at a relatively low welfare cost compared to alternative revenue sources. Ignoring disparities in "hassle costs" between the interventions, our estimate shows that the encouragement letter generates public revenue at a slightly lower welfare cost than audits do. However, we argue that this observation does not necessarily imply that the tax administration should prioritize the letter intervention in a scale-up.

---

[1]They also document some worrying signs about the literature to date which suggest selective reporting of results, in particular they find that larger studies tend to have smaller effects and that marginally significant effects are more likely to be reported than marginally insignificant effects.

2 Institutional Background, Data and Experimental Design

2.1 Tax filing timeline

Our design is closely linked to the sequence of actions and the information exchange between the Norwegian Tax Administration (NTA), third-party institutions, and personal taxpayers, see Table 1 for a detailed timeline of tax returns. Employers report employee earnings to the NTA and withhold stipulated taxes. Other individual income sources, e.g., interests and financial capital gains, are also reported by third parties, as are some deductions, including interest paid on mortgages and donations to charitable organizations. Based on third-party information, tax returns for year t are prefilled and distributed by the NTA to taxpayers at the beginning of April in year $t + 1$. Employees and pensioners can then make corrections to their tax returns and self-report income and/or deductions until April 30, while self-employed must file their personal tax report before the end of May.[2]

Our study focuses on self-reported deductions. The most common self-reported income tax deduction items are interest on debt, personal work-related expenses on costs related to stays away from home, childcare deductions, and expenses from lending out property (Løyland et al., 2024).

---

[2]Over the next months, and actually up to three years under the current tax law, personal taxpayers can reopen the file and adjust their reported items.

Table 1: **Timeline 2018. Personal Tax Returns for Tax Year 2017.**

	Standard procedures (business as usual)	**Actors**	**Field Experiment Treatments**	**Outcome short run**
January-February	Third party reporting	Employers and Financial Institutions		Income, interests, wealth
March	Pre-filled tax returns distributed	Norwegian Tax Administration (NTA)		Income by source, deductions, gross wealth, debt
April	Check, correct and self-report if relevant	Employees and pensioners		Acceptance of pre-filled or *self-reported* deductions and income
May	Check, correct and self-report if relevant	Self-employed		Acceptance of pre-filled or *self-reported* deductions and income
May-October	Programmed audit routines (flags)	NTA to taxpayers	**Letter (L=1)**	*Self-adjustment* by taxpayers
	Programmed audit routines (flags)	NTA	**Audit (A=1)**	Approval or *audit-adjustment* by the NTA
	Programmed audit routines (flags)	NTA	**Non-treatment (A=L=0)**	Approval or audit-adjustment by the NTA
October-December	Final assessment	NTA		*Final total deductions*, taxable income and wealth

Tax audits are carried out during the May–December year $t + 1$. Since 2014, two main types of audit have been used to verify self-reported itemized tax deductions of personal taxpayers. First, a traditional targeted audit is based on computer-generated flags that pop up if there are irregularities on specific items. The second type of audit is based on a broader set of information where every taxpayer

is given a risk score based on individual characteristics, recent filing, and historical records.

In 2013, the NTA singled out 310 000 taxpayers claiming self-reported deductions above an (unofficial) threshold of Z Norwegian kroner on one or two items from a list of 29 specified expenses. A random sample was checked to train and test a gradient-boosting machine learning algorithm to predict a binary classifier of compliance/noncompliance. In 2014-2016, the model provided a risk score for every taxpayer and those with a risk score above a specific year threshold were selected for audit, (Løyland et al., 2024).

2.2 The field study implementation

For the tax year 2017, the NTA used the prediction model to calculate a risk score for all personal taxpayers, based on their filed report by the end of April 2018 (end of May for the self-employed). From this distribution, around 15 000 individuals with the highest score were selected to form the target group. This high-risk filer population was randomly divided into three groups of equal size; two treatment and a control group.

One-third was drawn for a standard low-cost office-based audit. The audit checked for suspicious itemized tax deductions and asked for documentation from the taxpayer if necessary.[3] The taxpayer was only notified if the auditor found irregularities with the claimed deductions or asked the taxpayer for additional documentation. Hence, all taxpayers who had their deductions adjusted by the NTA knew they had been audited, but we do not have exact information on whether the compliant taxpayers knew they were audited. The auditors only checked whether the deductions were legitimate and ignored other items, including income.

The other treatment group received a softer intervention; a letter encouraging them to reconsider and check their self-reported itemized deductions. The letter was sent to the taxpayers between May and October 2018. The letter asks taxpayers to take a second look at their itemized self-reported deductions and states that "Random checks performed by the NTA show that 6 out of 10 taxpayers in your situation make mistakes when claiming this kind of deductions" (Appendix Figure A.1). It also reminds taxpayers that documentation must be provided upon request. The letter was sent through an electronic personal information platform used by the Norwegian authorities ("Altinn.no"). The taxpayers were notified once by an e-mail or text message that there is a letter from the NTA in their

---

[3]The NTA also run firm audits. Using data from randomly assigned on-site audits among 2 462 Norwegian firms, Bjørneby et al. (2018) provide evidence of collusive tax evasion whereby employers and employees collude to keep transactions off the books.

personal inbox. It is important to note that although the letter came after the tax report was filed, it is possible for taxpayers to reopen their report and check for mistakes and make adjustments. This is what the letter encourages.

The final third had their tax reports checked by standard procedures. Since regular desk audits of suspicious filing carried on as usual irrespective of this RCT, a small minority of the control group and the letter group experienced a flag audit. The taxpayers were not given any information about the audit selection mechanism, and since they followed the same protocol, we assume equal compliance effects of the two types of audits. We can therefore estimate the effect of audit, using the RCT assignment as an instrument. Moreover, a small fraction of the taxpayers selected for the letter treatment did not actually receive the message. The policy-relevant treatment is sending a letter (intention to treat), but the effectiveness of this enforcement depends on the extent to which the message is received, which in turn is affected by the delivery of technology and individual effort. In the results section, we discuss and present several ways of dealing with cross-overs, all of which lead to even larger behavioral effects of actually receiving the treatments.

3   Data

The data contains detailed tax filing records, as well as core demographic information of each tax payer. The main outcomes are deductions, i.e., different types of expenditures that can be subtracted to reduce taxable income. The field experiment was implemented on information initially submitted during the Spring 2018 tax return for the income year 2017. The three core items are (i) self-reported deduction before the intervention, (ii) the adjustments in the intervention year (by the taxpayer herself or the NTA auditor), and (iii) self-reported deductions next year.

3.1   Samples

Data on the initial submission of tax returns for the income year 2017 were extracted from the NTA data warehouse as of May 17 2018. The gross sample counts 14 902 with an equal share for the treatment groups (Table 2). Data were missing for 826 taxpayers because they had not yet submitted their tax return (mainly self-employed taxpayers). Tax return data were also missing or incomplete for another 1 108 taxpayers at this date. For these taxpayers, the main reason for missing data was that they had submitted their tax return on paper (non-electronically). In such cases, the NTA manually enters the tax return data into the tax systems, and this was done after the extraction date (17 May

6

2018). Due to the overwriting of previous versions of tax returns in the data warehouse, initial tax return data on these 1 934 taxpayers are not available.

Taxpayer filing data contain outliers stemming from different sources of measurement error and/or extreme random numbers. Previous studies have used different strategies to deal with these problems. While DeBacker et al. (2018) winsorize at p90, Kleven et al. (2011) trim income changes (after treatment) at −200 000 and +200 000 kroner "to eliminate extreme observations that make estimates imprecise" and Advani et al. (2021) "trim the top 10% to avoid outliers having an undue impact on the results". We follow the trimming practice, and exclude taxpayers with values above the 99th percentile for one or more of the four variables; prefilled deductions, pre-treatment claimed deductions, post-treatment claimed deductions and post-treatment final deductions. Our net sample for the analyses of short-run treatment effects therefore consists of 12 459 taxpayers. The sample reduction is similar for all treatment groups as shown in Table 2.

Turning to the compliance effects sample in the year following the treatments, there are missing values and attrition due to trimming, deaths, and migration. In particular, there are a large number of migrants in our initial sample and many of them are likely to have left Norway. While the decision to stay may be affected by the treatment, we do not view this as very likely. We tested and rejected that there is a difference between the groups in the probability of being present in the tax register. We further follow the same trimming practice as for the gross sample discussed above. The attrition from the short run to the compliance effect sample is just about 5 %.

Table 2: **Gross and net samples, attrition and sample exclusion**.

Criteria	Observations				Comments
	Audit	Letter	No	All	
Gross sample	4 964	4 945	4 993	14 902	
- Missing data due to late subs	285	285	256	826	Mostly self-employed
- Missing other reasons	376	366	366	1108	Manual submission, delayed handling by NTA
- Trimming	151	164	193	409	Taxpayers with deductions (4 items) above p99
Short run effect sample	4 151	4 130	4 178	12 459	
-No info 2018	52	49	43	144	Not present in the NTA tax liability register in 2019
-Technical attrition	104	109	94	307	System changes in the NTA
-Trimming	77	82	79	238	Taxpayers with deductions above p99
Compliance effect sample	3 918	3 890	3 962	11 770	

Since the target group is high-risk filers, they are not representative for Norwegian taxpayers. In Table 3, we present pre-treatment characteristics of the target group as well as for the whole population of

Table 3: **Pre-audit characteristics. Target population and all Norwegian taxpayers.**

| | Target population | | All tax payers | |
Variables	Mean	Std. Dev.	Mean	Std. Dev.
Income and deductions				
Pre-filled deductions	128 384	49 058	106 566	59 299
Self-reported deductions	67 690	42 966	4 431	23 619
Taxable income	315 878	244 654	311 920	335 387
Individual characteristics				
Age	40	11	47	19
Female	0.27		0.49	
Immigrant	0.38		0.06	
Married	0.26		0.42	
Risk score	0.7	0.08	n.a.	n.a.
Observations	12 459		4 445 228	

Note: All taxpayers include those 18 years and older. Variables in NOK (income and deduction variables) are trimmed such that observations above the 99th percentile are dropped. All monetary variables are in 2017-prices.

Norwegian taxpayers 18 years of age and older. We note that the high-risk filers have higher income, considerably higher self-reported deductions, they are younger and male-dominated compared to all taxpayers.

3.2 Outcomes

The short-run outcomes are adjustments of self-reported deduction. For the audit treatment, the adjustment is made by the NTA auditor. For the letter, the taxpayers themselves corrected the deductions.

The compliance effects in the next-year tax filing are measured by self-reported deductions as well as total claimed deductions. Total deductions include pre-filled third party reports and self-reported deductions and will therefore capture the behavioral effects of the enforcement treatments. If there is no effect of the enforcement treatments, the total claimed and final deductions are identical in the intervention year.

Unfortunately, we cannot use the field experiment to identify long-term behavorial effects, as the NTA laywers insisted on equal treatment, which implied a postponed audit for the letter and control group.

These audits were done in July-October 2019 and, therefore, affected the final deductions for the tax year of 2018. However, as the postponed audit for the letter treatment group took place after the initial filing for 2018 taxes, they did not affect self-reported deductions for 2018 which is our main outcome of interest.

3.3   Pre-treatment balance

We test for balance on a relevant set of pre-treatment variables. Most importantly, we did not find significant differences between the treatment groups in self-reported deductions (Table 4). The same holds for prefilled deductions by third parties and the tax authorities.

In Table 4, we test whether there are systematic differences between the audit group and the control group and then between the letter group and the control group. We test each variable separately and then conduct an F test of whether the variables jointly predict treatment status. The F-tests are passed for both treatments, and the individual variables seem balanced across groups. Overall, we found no indication that randomization failed to provide groups with expected counterfactual outcomes.

Table 4: **Pre-treatment balance.**

Characteristic	Audit	Letter	No	Audit vs No	Letter vs No
		Mean (std dev)		t-test p-value	
Women	0.27	0.28	0.27	0.819	0.374
Age	39.8	39.9	39.9	0.921	0.696
	(10.6)	(10.7)	(10.6)		
Married	0.27	0.26	0.26	0.612	0.807
Norwegian citizen	0.62	0.52	0.61	0.595	0.476
Self employed	0.04	0.04	0.04	0.770	0.897
Risk-score 2017	0.697	0.694	0.694	0.078	0.972
	(0.079)	(0.0797)	(0.079)		
Final total deductions 2015	140 636	140 460	140 742	0.949	0.863
	(70 582)	(69 416)	(71 413)		
Final total deductions 2016	144 398	144 324	144 269	0.929	0.969
	(62 833)	(62 420)	(63 125)		
Pre-treatment deduction 2017	195 893	195 649	196 674	0.555	0.437
	(60 878)	(60 282)	(59 961)		
Pre-filled deductions 2017	129 270	127 569	128 310	0.376	0.489
	(49 467	(48 246)	(49 441)		
Self-reported deduction 2017	66 623	68 080	68 365	0.063	0.765
	(42 303)	(43 325)	(43 251)		
Observations	4 151	4 130	4 178		
All variables				F-test	F-test
				F(10,8315)=0.60	F(10,8294)=0.36
				p-value = 0.81	p-value = 0.96

Note: Short-run effects sample; see Appendix for the future compliance sample. In the F-test we code eventual missing observations as zero and include a dummy variable for missing status in order not to lose observations.

4 Empirical model and predictions

To find the average compliance effects of the two tax enforcement policies, we estimate the following regression using ordinary least squares:

$$Y_{i,t} = a + b_t Letter_{i,t_0} + c_t Audit_{i,t_0} + d_t X_{i,t_0} + u_{i,t} \qquad (1)$$

We estimate the effects on deductions $(Y_{i,t})$ in the year of the interventions $(tax\ year = t = t_0 = 2017)$ and the year after $(t = t_0 + 1)$. To gain precision, we also include pre-treatment deductions as controls (X_{i,t_0}).

10

Taxpayers had already filed their tax report for the income year 2017 when exposed to either the letter or the audit treatment. However, both treatments can alter the final deductions for year $t_0 = 2017$. Although the letter encouraged the taxpayer to reopen their report and adjust self-reported tax deductions, the audit would lead to an adjustment by the NTA in case of any irregularity. We expect this short-run adjustment effect to be largest for the audit ($c_{t_0} < b_{t_0}$).

There are reasons to be genuinely uncertain about the short-term effect of the letter (b_{t_0}). We know that information hinting at increased deterrence tends to have a stronger effect on tax compliance than letters appealing to tax morality or civic duty (Slemrod, 2019). The letter is fairly neutral, without moral suasion or explicit statements that should increase their perceived detection probability. There is no explicit mention of injunctive or descriptive norms and it does not openly threaten that itemized deductions will be audited unless action is taken. Notwithstanding, those who receive the letter can interpret it as a signal that they are on the radar of the tax authorities. Of course, the encouragement also induces taxpayers to take a second look and correct unintentional mistakes. Among the taxpayers who want to pay their due taxes, it is reasonable to assume that a fraction have mistakenly filed too high deductions, and the letter motivates them to check the rules more thoroughly. We expect that a fraction of those with letter treatment will reopen their files and self-adjust their report simply because they were wrong. However, as long as not every taxpayer with irregular tax deductions self-adjusts, the average adjustment of those receiving a letter will be lower than among the audited taxpayers.

Regarding future compliance effects, existing empirical evidence makes us expect that audit exposure will lead to lower self-reported itemized tax deductions in the future ($c_{t_0+1} < 0$). However, there are potential mechanisms that may contribute to higher deductions. First, some lab evidence suggests that an audit today can reduce the perceived future audit probability (a "bomb-crater effect") (Mittone et al., 2017). Moreover, even if the risk of being audited is adjusted upwards, the assessed probability that noncompliance will be detected may go down for those audited without consequence (Gemmell and Ratto, 2012). Finally, the audits can also lower future compliance by showcasing that the penalties for noncompliance are low. For the letter, we see no role for these mechanisms and expect increased compliance ($b_{t_0+1} < 0$). It is not clear a priori which policy will have the largest effect. The audit adjustment is more intrusive and forceful, but some of the audited will not know that the tax administration has checked their files because they were not adjusted or asked for documentation. In contrast, the letter was sent to all assigned to this treatment.

5 Results

Our main results are reported in Table 5, columns (4) and (5). The short run average treatment effects are reported in Panel A as entries from separate linear OLS regressions, controlling for pre-treatment claimed deductions.

The audits disclosed extensive illegitimate deductions. Two in three taxpayers (65.3%) had their reported deductions corrected by the auditor. Among the taxpayers exposed to business as usual, only 6% were adjusted via the ordinary flag-based audit. Therefore, the audit treatment increased the fraction who had their deductions adjusted downwards by 59 percentage points. The average audit adjustment -29 538 NOK (1 USD= 8.3 NOK in 2017), or 43% of the average self-reported deductions. Among those adjusted, on average 50 064 NOK of the deductions was not approved by the NTA.

The letter had a significant impact on the filing behavior as shown in column (5) in Table 5. About 11% reopened their files and reduced their self-reported deductions. As close to none did in the business as usual group, the short-run letter effect is estimated to -0.105 and highly significant. When we combine the self-adjustment share of 0.105 with an audit hit rate of 0.653, the evidence suggests that about one in six taxpayers who had made a mistake did actually respond to the letter. The average self-adjustment effect of the letter is -3 584 NOK. The self-adjustment among those who responded (compliers) was nearly ten times larger and estimated to be -34,133 NOK. However, the effect of the letter on the final total deductions is smaller; -2 503 NOK. This is because 6% of the letter and control group had their deductions adjusted by the NTA through flag audit. Thus, a substantial part of the errors corrected by self-adjustment would have been discovered by standard procedures.[4]

While only the letter allows for any behavioral responses in the year of treatment, both interventions potentially affect future compliance in terms of self-reported deductions. In Panel B of Table 5 we report deductions for the following tax year, reported about ten months after the treatments. First, the pre-filled deductions from third parties are slightly lower for the two treatment groups, but there are no significant differences compared to the no treatment group. Turning to the treatment effects on taxpayer self-reported deductions, we see that both interventions reduced the fraction with self-reported deductions. The audit lowered the share with nontrivial self-reported deductions by 12 percentage points (pp). The letter effect is lower (5 pp), but statistically and economically significant.

---

[4]For final total deductions, the audit adjustment by far exceeds that the self-adjustment from the letter (-30 159 NOK vs -2 503 NOK).

Table 5: **Compliance effects.**

	Sample means			Regression coefficients	
	Audit	Letter	No	Audit	Letter
	(1)	(2)	(3)	(4)	(5)
Panel A. Short run					
Share with self-adjustment	0.014	0.110	0.005		0.105***
					(0.005)
Self-adjustment	-135	- 3 084	474		-3 584***
					(293)
Share audit adjusted	0.653	0.062	0.061	0.59***	
				(0.008)	
Audit adjustment	-34 071	-3 845	-4 674	- 29 538***	
				(731)	
Final total deductions	161 687	189 079	192 474	- 30 159***	-2 503***
				(738)	(573)
Panel B. Future compliance					
Share self-report ded. > 1 000 NOK	0.539	0.610	0.660	- 0.12***	- 0.05***
				(0.01)	(0.01)
Self-reported deductions	25 155	31 382	35 283	-10 093***	-3 722***
				(898)	(898)
Claimed deductions	161 909	167 814	172 883	-10 825***	-4 583***
				(1 351)	(1 094)
Sample sizes					
Panel A	4 151	4 130	4 178	8 329	8 308
Panel B	3 918	3 890	3 962	7 880	7 852

Note: All numbers are in NOK, except for fractions. Panel A (B) uses outcomes for the tax year 2017 (2018). In columns (4) and (5), each entry is from a separate linear OLS regression and reports the treatment dummy estimate with standard error in parenthesis. Total pre-treatment deductions is included as control in the regressions. *** : significance at the 1%-level.

The effects on self-reported deductions are also significant.[5] The audit effect is -10 128 NOK, or 29% of the self-reported deductions in the no-treatment group. The letter effect is smaller (- 3 900 NOK), but also statistically significant. Even if the future compliance effect is considerably larger for audits, they are more similar to the short-run adjustment effects.

Taxpayers with a spouse will typically not make filing decisions in isolation. Some deductions are specific to the household and can potentially be transferred from one spouse to the other as a response to treatment. Spouses may also update their knowledge about tax rules and audit probabilities when their partner has been subjected to the audit or the letter. Both mechanisms suggest that spousal reporting is part of future compliance effects. Estimates based on household outcomes are very similar to the individual effects reported in panel B of Table 5.

While the interventions focus on self-reported deductions, taxpayers may adjust their income and/or how they report it in response to the audit or letter. In Table A.2, we reestimate the models with taxable income as the outcome. For audits, the short-run and compliance effects estimates are very similar (with opposite sign) to those for deductions in Table 5. This rules out any effects of audit on total income before deduction. For the letter, the estimated effects are imprecisely estimated and not significantly different from zero. Transitory components are more important for income than for deductions. For taxable net income, which includes items from third parties and self-reports, the standard deviation is even greater. Given the large variability of income, we do not have statistical power to detect effects of the letter on taxable income.

As in most RCTs, some participants received a treatment different from the one to which they were randomly assigned.[6] First, a minority of the non-treatment group experienced a flag audit. The flag and treatment audits followed the same protocol. Since taxpayers did not know why they were selected (risk score threshold or single items with a flag), we assume that they have the same behavioral effects on future compliance. This motivates the assigned treatment audit as an instrument for any audit. This basically scales the effect of audit in Table 5 by the inverse of the increase in the share with audit due to the random assignment. About 22% of the nontreatment group were audited due to flags. Therefore, the IV estimate of the actual audit in Table 6 is somewhat larger (in absolute numbers) than the effect of assigned audit in Table 5.

---

[5]The self-reported deductions are clearly lower than in the previous year for all three groups. This mean reversion reminds us that treatment effects are hard to identify from data based on operational audits triggered by "suspiciousreporting ".

[6]This is typically called non-compliance, but this label has another meaning in this paper.

Table 6: **IV estimates accounting for cross overs.**

	Audit	Letter (sent)	Letter (sent and opened)
Short run (Panel A of Table 5):			
Self-adjustment (NOK)	not relevant	-3 751***	-3 871***
		(306)	(316)
Audit adjustment (NOK)	-37 750***	not relevant	not relevant
	(965)		
Sample size	8 329	8 308	8 308
Future compliance (Panel B of Table 5):			
Self-reported deductions (NOK)	-12 949***	-3 918***	-4 037***
	(1 140)	(932)	(962)
Claimed deductions(NOK)	-13 887***	-4 760***	-4 905***
	(1 396)	(1 137)	(1 172)
Sample sizes	7 880	7 852	7 852

Second, the letter was not sent to all assigned taxpayers due to some failing administrative procedures. For about 3.7%, no letter was sent. We can estimate the effect of the letter sent by instrumenting with the assigned letter and find that the compliance effect estimate increases slightly. The effect of a sent letter also depends on the share of taxpayers who actually receives the message. From a behavioral insight perspective, we would like to know the average effect of receiving the message of the letter. Again, we can use letter assignment as an instrument for sent and opened. Since 93.4% received the message, the effect of receiving the message (last column of Table 6) is slightly larger than the intention-to-treat estimate in Table 5.

6 Welfare analysis

Table 7 reports the costs and tax revenues from the letter and audit. Tax revenue is found by multiplying the compliance effect, that is, the estimated reduction in self-reported income tax deductions caused by the interventions, by the relevant tax rate at that time, which was 23%. According to NTA estimates, it costs 1 625 NOK to conduct this type of audit and the letter had a unit cost of 266 NOK. These costs include wages and social costs, e.g. pay roll taxes, mandatory employer insurance and holiday pay, IT-equipment costs, and office costs. Both interventions collect considerable more tax revenue than their costs.

What criteria should we use to prioritize these interventions? If we decide to implement one of the ini-

tiatives on a broader scale for the entire high-risk filer group, should it be the letter or the audit? On a more fundamental level, should the government explore alternative methods to increase tax revenue instead of relying on these tax enforcement strategies?

Table 7: **Net tax revenue effect (in NOK).**

	Audit	Letter
a. Short run deductions	-29 538	-3 584
b. Next year deductions	-10 128	-3 900
c. Sum deductions $(a + b)$	-39 666	-7 484
d. Tax revenue $(0.23 * c)$	9 123	1 721
e. Unit costs	1 625	266
Net tax revenue $(d - e)$	7 494	1 455
MVPF $(d/(d - e))$	1.22	1.18

Governments can collect more revenue either by increasing tax rates or by broadening the tax base, for example through intensified tax enforcement to enhance compliance. The theoretical literature on tax administration explores the optimal use of both of these tools. The fundamental principle that guides optimal tax administration is that the welfare cost per dollar of tax revenue from a marginal increase in tax enforcement should be equal to the welfare cost per dollar of revenue obtained through a marginal increase in the tax rate (Keen and Slemrod, 2017). Boning et al. (2023) argue that a practical approach to implementing this principle is to compare the marginal value of public funds (MVPF) generated by increased tax enforcement with the MVPF resulting from a slight increase in the tax rate.

It is beyond the scope of this paper to make a comprehensive comparison of the welfare effects of an increase in the income tax rate with an increase in tax enforcement. However, we can estimate the MVPF for each of the two enforcement policies that we have studied. For tax revenue interventions, MVPF is defined as the private willingness to pay to avoid the intervention divided by the net government revenue collected by the intervention.

$$MVPF = \frac{\text{Willingness to pay to avoid intervention}}{\text{Net increase in government revenue from intervention.}}$$

The nominator captures what individuals are willing to pay not to have the intervention implemented. In a tax-enforcement setting, this amount obviously includes the additional taxes they have to pay,

but it also include penalty taxes and the monetary value of the hassle for taxpayers associated with the enforcement policy. Since only a very small fraction of the corrected taxpayers have to pay a penalty(Hebous et al., 2023) and since we have no estimate of the "hassle" costs, we ignore these terms in the calculations below. The denominator of MVPF is simply the revenue effect of the tax enforcement policy.

Policies to raise public revenue are better the lower MVPF is. A MVPF equal to one implies that there is no efficiency loss associated with this revenue source. Table 7 shows that, according to the MVPF criteria, the letter is ranked at or slightly above the audit. When doing the calculations, we obtain $MVPF^{letter} = 1.18$ and $MVPF^{audit} = 1.22$. This means that the letter collects one unit of revenue at the minimal welfare cost. However, it does not automatically imply that the NTA should scale up and apply the letter to all high-risk filers. The decision hinges on the specific objectives set by the tax administration.

If the goal is to collect a fixed revenue at the lowest welfare cost, the letter is the best alternative. If the goal is to collect maximal net tax revenue from this group of high-risk filers, NTA should implement the audit. In this particular case where the $MVPF's$ are almost identical, the natural choice seems to be to scale up the audit and collect a substantial higher amount of tax revenue at a slightly higher welfare cost per revenue unit.

A critical caveat when calculating the welfare implications of letters or other information interventions is that it remains uncertain whether similar interventions in subsequent years would yield consistent effects. This uncertainty relates to a phenomenon known as 'nudge fatigue,' which occurs when individuals become desensitized or resistant to the influence of repeated interventions.

7 Conclusion

In a large-scale tax enforcement field experiment, we find that two alternative policies significantly affected the filing of illegitimate self-reported tax deductions by personal taxpayers, both in the year of intervention and for the subsequent tax year. The desk-based correspondence audit had a larger effect than the letter encouraging tax filers to take a second look at their tax deductions. Even if audits are more costly, they generate higher net tax revenue than sending a letter. If we alternatively use the marginal value of public funds to rank these policies, sending an encouragement letter is at par with audit as the best intervention.

References

Advani, A., W. Elming, and J. Shaw (2021). The dynamic effects of tax audits. *The Review of Economics and Statistics*, 1–45.

Alm, J. (2019). What motivates tax compliance? *Journal of Economic Surveys 33*(2), 353–388.

Antinyan, A. and Z. Asatryan (2019). Nudging for tax compliance: A meta-analysis. *ZEW-Centre for European Economic Research Discussion Paper* (19-055).

Bergolo, M. L., M. Leites, R. Perez-Truglia, and M. Strehl (2020). What makes a tax evader? Technical report, National Bureau of Economic Research.

Bjørneby, M., A. Alstadsæter, and T. Kjetil (2018). Collusive tax evasion by employers and employees: Evidence from a randomized field experiment in norway.

Boning, W. C., N. Hendren, B. Sprung-Keyser, and E. Stuart (2023). A welfare analysis of tax audits across the income distribution. Technical report, National Bureau of Economic Research.

Bott, K. M., A. W. Cappelen, E. Ø. Sørensen, and B. Tungodden (2020). You've got mail: A randomized field experiment on tax evasion. *Management Science 66*(7), 2801–2819.

De Neve, J.-E., C. Imbert, J. Spinnewijn, T. Tsankova, and M. Luts (2021). How to improve tax compliance? evidence from population-wide experiments in belgium. *Journal of Political Economy 129*(5), 1425–1463.

DeBacker, J., B. T. Heim, A. Tran, and A. Yuskavage (2018). Once bitten, twice shy? The lasting impact of enforcement on tax compliance. *The Journal of Law and Economics 61*(1), 1–35.

Fellner, G., R. Sausgruber, and C. Traxler (2013). Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association 11*(3), 634–660.

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

Hallsworth, M. (2014). The use of field experiments to increase tax compliance. *Oxford Review of Economic Policy 30*(4), 658–679.

Hebous, S., Z. Jia, K. Løyland, T. O. Thoresen, and A. Øvrum (2023). Do audits improve future tax compliance in the absence of penalties? evidence from random audits in norway. *Journal of Economic Behavior & Organization 207*, 305–326.

Holz, J. E., J. A. List, A. Zentner, M. Cardoza, and J. E. Zentner (2023). The 100 million nudge: Increasing tax compliance of firms using a natural field experiment. *Journal of Public Economics 218*, 104779.

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

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), 651–692.

Løyland, K., O. Raaum, G. Torsvik, and A. Øvrum (2024). Evaluating compliance gains of expanding tax enforcement. *Economica*.

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*, 225–243.

OECD (2019). *Tax Administration 2019: Comparative Information on OECD and other Advanced and Emerging Economies.* OECD Publishing, Paris.

Pomeranz, D. and J. Vila-Belda (2019). Taking state-capacity research to the field: Insights from collaborations with tax authorities. *Annual Review of Economics 11*, 755–781.

Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature 57*(4), 904–54.

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

A   Tables and figures referred to in the text

	Return address:	Our date	Your date	Executive officer
	P O Box 6499 Etterstad, N-0606 OSLO		Your reference	Telephone
Skatteetaten			Our reference	Postal address

Name
Address
Postcode and town

Check your deductions

In your tax return for the 2017 income year, you have claimed a deduction under item
- x.x.x. (item name/deduction)
-

From the 2016 income year, you're required to assess the basis for your tax calculation yourself. Random checks performed by the Norwegian Tax Administration show that 6 out of 10 taxpayers in your situation make mistakes when claiming this kind of deduction.

What do I have to do?
You have to check that you've given us the right information, and that you can document the deductions you've claimed in your tax return.

You have to enter the correct information in your tax return. If you find errors, correct them and submit your tax return at https://www.skatteetaten.no/en/person.

If we ask you to, you must be able to show documentation (receipts etc.) for the deductions in your tax return.

Please disregard this letter if the amount entered in your tax return is correct.

Do you have any questions?
You can read more about the items at https://www.skatteetaten.no/en/person/taxes/tax-return/find-item/. Call us on 800 80 000 if you have any questions.

Yours sincerely,

Figure A.1: Letter to taxpayer. Check deductions

Table A.1: **Compliance effects sample. Balance across treatment assignments. Means, standard deviation and test.**

Characteristic	Audit	Letter	No	Audit vs No	Letter vs No
		Mean (std dev)		t-test p-value	
Women	0.276	0.280	0.273	0.741	0.450
Age	40.0	40.0	39.8	0.587	0.424
	(10.7)	(10.7)	(10.5)		
Married	0.272	0.269	0.265	0.481	0.653
Norwegian citizen	0.617	0.618	0.608	0.406	0.377
Self employed	0.037	0.037	0.037	0.888	0.842
Risk-score 2017	0.697	0.694	0.694	0.168	0.931
	(0.079)	(0.077)	(0.080)		
Final total deductions 2015	140 667	139 668	140 029	0.701	0.826
	(69 026)	(67 660)	(70 183)		
Final total deductions 2016	144 439	144 211	144 094	0.810	0.935
	(61 008)	(60 892)	(61 853)		
Pre-treatment deduction 2017	195 107	194 561	195 355	0.851	0.548
	(59 320)	(58 720)	(58 331)		
Pre-filled deductions 2017	128 591	126 607	127 151	0.175	0.080
	(47 218	(45 442)	(47 057)		
Self-reported deduction 2017	66 515	67 954	68 204	0.602	0.798
	(42 297)	(43 424)	(43 174)		
Observations	3 918	3 890	3 962		
All variables				F-test	F-test
				F(10, 7 866) = 0.54	F(10, 7 838) = 0.37
				p-value = 0.86	p-value = 0.96

Note: All numbers are in NOK, except for fractions. Each entry is from a separate linear OLS regression and reports the treatment dummy estimate with standard error in parenthesis. The pre-treatment taxable income is included as control in the regressions. *** indicates significance at the 1%-level.

Table A.2: **Effects on taxable income.**

	Audit	Letter
A. Short run	31 775***	3 423
	(1 996)	(2149)
B. Future compliance	8 829***	749
	(3 582)	(3 618)

Note All numbers are in NOK.