A Pre-Analysis Plan for "Do Think Twice, it´s All Right: Effects and Mechanisms of Tax Enforcement Policies"

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

3 October 2019

Abstract

We compare the tax compliance effects of two different tax enforcement policies, a relatively costly desk/correspondence audit and a cheap information treatment with a letter asking filers with "risky" filing behavior to take a second look at their tax returns. To assess the total effect of these enforcement policies, we need to take account of both their immediate effects and their long term effects on future filing behaviour. Audits will detect and correct non-compliance on the spot, but may also change subsequent filing behavior. A letter will not disclose non-compliance, but (some) taxpayers may adjust their filings. A letter may also have long term effects on tax compliance. Based on the short run effects of both treatments, which are easy to measure, we describe how we will test the behavioural effects on future filing. We describe the intervention and lay out some important decisions with respect to coding of variables, definitions of samples and the empirical strategy we will apply.

1 Introduction

It is important to ensure that taxpayers follow the tax rules and pay their due taxes, since non-compliance creates both efficiency and fairness losses. To enforce compliance, the tax authorities carry out audits. Audits are costly and targeted towards individuals with indications of risky filing behavior. It has been demonstrated in several studies that in addition to detect and correct infringement on the spot, audits may have lasting effects on taxpayers subsequent compliance with the tax code, [Advani et al., 2017, DeBacker et al., 2015, Løyland et al., 2019]. A softer, and less costly tax enforcement strategy than audits, is to send taxpayers with high risk profile a letter encouraging them to take a second look at their files to see if they have done a mistake. Such a letter may have both immediate effects, as some taxpayers may choose to self-correct this years tax files, as well as long term effects on tax compliance in subsequent years.

In this study we work together with the Norwegian National Tax Authorities (NTA) to compare the effectiveness of these two enforcement policies; a desk based correspondence audit and a "take a second look" letter, referred to as the Audit and the Letter hereafter. We expect the Letter to have a more moderate effect on tax compliance than the Audit. From a cost-benefit perspective, however, the Letter can be the more efficient policy since the costs are negligible compared to the Audit.

We randomize the Letter and the Audit in a population of nearly 15 000 personal taxpayers who claimed relatively high self-reported income tax deductions. A random third of this population got the Letter, another third were exposed to the Audit, and one third where not exposed to any intervention. We will consider both the effect on the treated taxpayers and any spillovers to the spouse.

In order to get a better understanding of the mechanism behind potential behavioural responses to the interventions, we conduct a survey to assess how and whether the interventions influenced taxpayers perceived probability of being audited in the future. According to standard theory, an increase in the perceived probability that evasion will be detected will raise compliance. It is, however, debated how an audit affect the perceived probability of future audits, and also how an audit affects the likelihood that evasion is detected if they are audited [Gemmell and Ratto, 2012, Mittone et al., 2017]. For the Letter on the other hand it is hard imagine that it can lower the recipients perceived probability of future audits.

The structure of this pre-analysis plan is as follows. Section 2-4 describes the institutional details, the treatments, samples and short run effects from data already accessed. From section 5 onwards, we lay out strategies for estimating subsequent compliance effects and tests of mechanisms based on data that will be available to us after the filing of this plan.

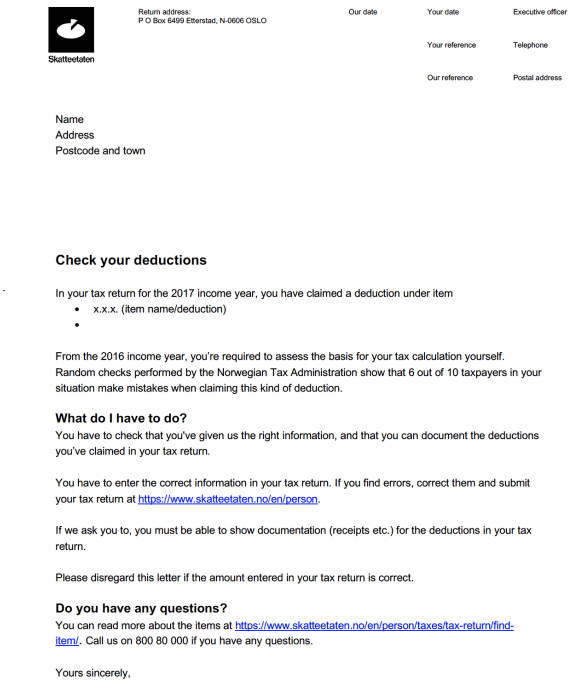
2 Treatments and survey

We consider two treatments. The enforcement interventions are conducted on a selected population of taxpayers with a risky profile, ie. where individual characteristics indicate a high likelihood that they are not complying with the tax code. We provide more details of this population in section 3.2.

One third of this population is selected (randomly) to be audited. The audits are standard low-cost office-based audits, commonly labeled correspondence audits [Hodge et al., 2015]. We define the treatment as being audited. In this type of audit, taxpayers are only notified that they are audited

it the auditor find some irregularities with the claimed deductions. Hence only those who obtain an adjustment can respond in subsequent years to being audited. It is also of some interest, although not directly policy relevant, to check the effects on future filings of those who obtained an adjustment of self reported deductions.

Another third received a letter encouraging them to reconsider and check whether their filed deductions were correct. The letter that was sent to the taxpayers after they self-reported deductions is presented in Figure 1.



Skatteetaten

Return address:
P O Box 6499 Etterstad, N-0606 OSLO

Our date | Your date | Executive officer
Your reference | Telephone
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 1. Letter to taxpayer. Check deductions

The taxpayer is asked to take a second look at the self-reported deductions and check if the filed information is correct. The letter also refers to evidence from previous audits where 6 of 10 were found to "make mistakes". Finally, the letter reminds the taxpayer that documentation must be provided upon request. The policy relevant treatment is defined as a letter being sent.

The letter was sent via an electronic personal information platform used by different Norwegian authorities. The taxpayer is notified by an e-mail or SMS that there is letter in the personal inbox from the NTA. There is only this one message, no reminder. It is likely that some of the recipient will not log on to the system and read the letter. Again it has some interest to also estimate the future filing effects of those who logged on the system and opened the letter.

The remaining one third of the sample is subject to business as usual. These taxpayers may obtain ordinary tax audits that are not part of our experiment. These audits are also correspondence audits, but audit selection is based simple rules, or "audit flags" (i.e., not risk scores). These audit flags are typically related to thresholds for self-reported data on specific income or deduction items. These other audits are independent of the random audit we consider, and will therefore not interfere

Table 1. Stylized Timeline of Employee Tax Returns for Tax Year t

Period Year $t+1$	**Action (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	Taxpayers		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 tax payers
	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

with our estimate of the future compliance effects of the examined audits.These are audits that are automatically generated if a tax payer display some filing behaviour.

3 Data structure

3.1 Tax filing and treatment timeline

Given our research question, it is important to have a clear understanding of the sequence of actions and the information exchange between the NTA and taxpayers. Table 1 details the timeline of tax returns for employees. As shown, the filing of tax returns occurs during April and May following the end of the income calendar year. Employers report taxable income to the NTA and they withhold the stipulated amount of taxes workers must pay. Other sources of individual income (such as capital income) are reported by third parties (including financial institutions). Some of the itemized tax income deductions (including donations to charitable organizations) are also reported by third parties (such as the receiving organization). Based on the third-party information, tax returns are pre-filled and distributed by the NTA to taxpayers at the beginning of April. Taxpayers can then make corrections to their tax returns and self-report income and/or deductions until April 30. Over the next months (actually up to three years under the current tax law), the taxpayer can self-adjusted their tax items.

Tax audits are carried out during the May–December period the following income year. Today there are two main types of tax audits for taxpayers. The first is the business as usual system that uses computer generated flags depending on some specific features of the tax return. The second type of tax audit is also targeted, but is randomly selected from the sample defined by the predictive machine learning models that produce taxpayer-specific risk scores. In both cases, the audits may not approve some deductions and make an audit-adjustment.

3.2 Population and samples

The sample consists of high-risk tax payers that score above a threshold of risk scores. The risk scores are based on a predictive machine learning model that produce taxpayer-specific risk scores using a large set of individual characteristics, including the tax return and taxpayer history.

The gross sample consists of 14 902 taxpayers (Table 2). To analyze the short run treatment effects (and future compliance effects), we need data from both the initially submitted tax return for income year 2017, and any self-adjustments by the taxpayer or audit adjustments by the NTA after this initial submission. Data on the initial submission of tax returns for the income year 2017 were extracted from the data warehouse of the NTA as of May 17 2018. At this date, tax return 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 were that they had submitted their tax return on paper (non-electronically). In such cases, tax return data are entered manually into the tax systems by the NTA, and this was done after the extraction date (May 17 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. Our net sample for the analyses of short run treatment effects (and future compliance effects) therefore consists of 12 968 taxpayers.

Filing data on taxpayers 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. [2015] winsorize at 90 percent, Kleven et al. [2011] trim income changes (post treatment) at $-200{,}000$ and $+200{,}000$ kroner "to get rid of extreme observations that make estimates imprecise" and Advani et al. [2017] "trim the top 1 per cent to avoid outliers having an undue impact on the results". We are following the practice of trimming, and exclude tax payers with values above the 99th percentile for one or more of the four variables; pre-filled deductions, pre-treatment claimed deductions, post-treatment claimed deductions og post-treatment final deductions.

3.3 Pre-treatment balance

We test for balance on a set of pre-treatment values of relevant variables. The results are seen in Table 3. We test whether there is a difference between the Audit group and the control group (Non-treat) and then between the Letter group and the control group. First we test each variable separately and then we conduct an F-test of whether the variables jointly can predict treatment status. The F-tests are passed for both treatments and the individual variables seem balanced across groups. 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.

Table 2. Gross and net samples, attrition and sample exclusion across groups

Criteria	Observations				Comments
	Audit	Letter	Non- treatment	All	
Total 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 99th percentile excluded
Final sample	4 151	4 130	4 178	12 459	For short run effects analysis and sample for compliance effects

Table 3. Balance across treatments: Means, standard deviation and test.

Characteristic	Audit	Letter	Non-treatment	Audit vs Non-treat	Letter vs Non-treat
	mean (std dev)			t-test p-value	
Women	0.27	0.28	0.27	0.819	0.374
Age	40.1	39.9	40.3	0.657	0.472
	(18.6)	(10.7)	(32.4)		
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)		
All variables				F-test	F-test
				F(9, 8316) = 0.48	F(9, 8294) = 0.26
				p-value = 0.88	p-value = 0.98

Table 4. Short run treatment effects. Deductions tax year 2017.

	Audit	Letter	Non-treat	Audit	Letter
		Means		Regression coeff vs Non-treatment (p-values)	
A. Pre-filled deductions	129 270	127 569	128 310	960 (0.376)	-741 (0.489)
B. Self-reported deductions	66 623	68 080	68 365	-1 741 (0.063)	-284 (0.765)
C. Claimed deductions ((A+B))	195 893	195 649	196 674	-781 (0.555)	- 1 025 (0.437)
D1. Share with self-adjusted*	0.014	0.110	0.005		0.105 (0.000)
D2. Self-adjustment (NOK)*	-135	3 084	474		-3 584 (0.000)
E1. Share audit adjusted*	0.653	0.062	0.061	0.592 (0.000)	
E2. Audit adjustment (NOK)*				- 29 538 (0.000)	
F. Final total deductions (C+D2+E2)*	161 687	189 079	192 474	- 30 159 (0.000))	-2 503 (0.000)
Sample sizes	4 151	4 130	4 178		

Note: * Total pre-treatment deductions as control in the regression. p-values in parentheses.

4 Short Run Treatment Effects

With short run treatment effects we mean effects in the year of the treatment. The taxpayers have already filed their report when they get either of the two treatments; a Letter or an Audit. Both treatments can alter the final tax files for that year. In the case of the Letter this happens if the recipient reopens the tax report that he or she has filed and corrects self-reported tax deductions. In the case of the Audit this happens if the auditor discovers non-compliance and adjusts the files.

We consider the short term effects on different outcomes specified in Table 4. The short run effects on these outcomes are average treatment effects, equal to differences in means between the treatment, either Letter or Audit, and the control group. To potentially gain precision we also report regression coefficients conditional on pre-treatment deductions.

Table 4 depicts that the share with self-adjustment is 11 percent for the letter group and close to zero for the non-treated. Conditional on pre-treatment deductions, the self-adjustment is about 3 584 NOK larger in the letter group. Obviously, there are tax payers who have over-reported deductions who do not respond to the letter, but are adjusted by the NTA in case of an audit. Presumably, few taxpayers will correct or withdraw deductions that are actually correct according to the tax rules.

Turning to the Audit, we first observe that the share who obtain audit adjusted taxes is 65.3 percent for those who are exposed to this audit, while it is only 6% for those who are in the control group (6% is also the number for the Letter group, indicating that in this population around 6% would obtain an ordinary flag audit, and hence that the effect of the audit we study is to increase the fraction who are audit adjusted by 59 percentage points). The audit group get an audit adjustment of -29,538 NOK as compared to the control group when controlling for pre-treatment deductions. In total we see that the effect of the audit is larger than the effect of the letter on the final total deductions.

Table 5. Outcome and controls. Coding and definitions.

Variable type	Content	Definition/description
Main outcome	Claimed deductions 2018 (t_0+1)	Pre-filled + self-reported (pre-audit year t_0+1). Trimmed
Secondary outcomes	Household Claimed deductions 2018 (t_0+1)	Aggregate Claimed deductions in a household
Controls	Claimed deductions 2017 (t_0)	Pre-treatment year t_0 (Trimmed, from Table 4)
	Age	Years since birth by 31 Dec
	Immigrant	Resident in Norway, foreign born with foreign born parents
	Temporary Labour Migrant	Non-resident, with D-number, citizenship
	Marital status	Dummy = 1 if married/cohabitat
	Risk score 2017	A continuous variable in $[0,1]$
Groups	Self employed	Dummy = 1 if self employed
	Female	Dummy = 1 if registered as woman

5 The effect on future tax compliance

5.1 Coding outcomes and controls

In this section we present the variables to be used in the analysis of future compliance. We start with the main outcome variables and continue with the covariates and the variables used to study heterogeneous effects. Our main outcome is the claimed deductions for tax year 2018 (in NOK). We cannot use final total deductions as an outcome (as we did for short run effects) since the field experiment was implemented in such a way that the group that received a letter in 2018 were audited in July-October 2019. Self-reported deductions were filed before this "extra audit" and are not affected by the 2019 audit, since the taxpayer had no information about the forthcoming audit.

We view the claimed deductions as a good outcome measure as it is highly influenced by behaviour in terms of self-reported deductions. In addition, the measure is equal to the final total deductions (and thereby net taxable income) in case of no audits. Our secondary outcome is the partner's claimed deductions for the tax year 2018 (in NOK).

The control variables and the codings are described in Table 5. These control variables will also be used to test for balance where we create a Table as the one in Table 4. The success of the randomization will be judged by the F-tests in the regressions where we include all the variables together and see whether they predict the treatments together. The control variables will also be used in exploratory analyses of heterogeneous treatment effects (see below). Whenever a control variable is included in a regression together with other variables we will code eventual missing observations as zero and include a dummy variable for missing status in order not to lose observations.

5.2 Main empirical specification and hypothesis

We estimate the following regression using ordinary least squares:

$$Y_{t_0+1,i} = a + bLetter_i + cAudit_i + u_i \tag{1}$$

$Y_{t_0+1,i}$ is the claimed deductions for tax year 2018 (in NOK) and we use robust standard errors. The main specification is a regression using individuals still present in Norway (unless there is non-random attrition, see below) without any controls included, if there is no imbalance across the groups.

If there is imbalance across the groups, the main specification will be one with the full set of control variables.

Our main hypotheses are

- (i) $b < 0$
- (ii) $c < 0$
- (iii) $b < c$

Taxpayers with a spouse will typically not make filing decisions in isolation. Some deductions are household specific and can potentially be transferred from one spouse to the other as a response to the Letter and Audit treatment. Spouses may also update their knowledge about tax rules or audit probabilities when their partner has been subjected to Audit or Letter. Both mechanisms make it important to include the filing behavior of the spouse in the estimation of future compliance effects of these treatments. We therefore also run the same regressions for the spouses.

We will also see if precision can be improved by adding the control variables described above. In particular, we will investigate if we can improve precision in the estimates by picking optimal controls from the total list of controls using a double debiased LASSO procedure [Belloni et al., 2014].

5.3 Test of mechanisms

In order to get a better understanding of the mechanism behind potential behavioural responses to the interventions we conduct a survey to assess how the interventions influenced taxpayers perceived probability of being audited in the future. In particular, we fielded a phone survey where we tried to contact all individuals in the main sample. We were not able to get the phone numbers for everyone and not everyone responded. Our main question of interest is Subjective Detection Risk (SDR) and is as follows:

"What do you think the probability is that the tax authorities will control your reported taxes in 2019?". The answer categories are

- 1) Not likely at all (0 percent).
- 2) Very unlikely (1-20 percent).
- 3) Quite unlikely (21-40 percent).
- 4) Neither likely nor unlikely (41-60 percent).
- 5) Quite likely (61-80 percent).
- 6) Very likely (81-99 percent).
- 7) Certainly (100 percent).

We will retain the continuous coding of this variable (1-7) and call it SDR and run the following OLS regression:

$$SDR_i = e + dLetter_i + fAudit_i + u_i \quad (2)$$

We expect that $d \geq 0$. We cannot think of any reasons why a letter with this content should induce recipients to update their perceived probability of an audit downwards. For the effect of Audit (f) there is a literature arguing that being audited today may reduce the perceived probability being picked out in future audits, see Mittone et al. [2017]. Hence the sign of f is a-priori uncertain, but we find it more likely that $f > 0$.

It is unclear what variables we will be able to connect to the survey from the register data for reasons of anonymity. The main specification will in any case be one without any control variables but the tests of balance as well as the heterogeneity analyses may be restricted to the subset of the variables we are able to include.

6 Sample, attrition and trimming

We use the "Gross sample" from Table 2 as our point of departure in analyzing the effects on future tax compliance. There will be missing values and attrition from this sample 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 will test whether there is a difference between the groups in the probability of being present in the tax register. The compliance effects will be tested for those present in the tax liability register of the NTA for 2019. If there is non-random selection into presence in the register we will also present estimates where we include everyone and code those not in the register as having zero deductions. We will further follow the trimming practice for restricting the gross sample as laid out in section 3.2.

With respect to the SDR mechanism, the sample will further be reduced by non-response to the survey. For that sample we will test attrition in the following way:

$$InSDRSample = g + hLetter_i + kAudit_i + u_i \quad (3)$$

Where *InSDRSample* is a dummy for being in the survey sample with valid information on question about detection probability in 5.3. If there are statistically significant coefficients for h and k (at the 5 percent level) we will follow Kling et al. [2007]'s correction. We will obtain lower bounds of the treatment effect by replacing missing observations in the treatment (control) arms by the corresponding arm's mean value minus (plus) 0.05, 0.10 and 0.20 standard deviations of the control group. Upper bounds of the treatment effects will be constructed in a symmetrical way.

7 Exploratory analyses

We will explore heterogeneity in the treatment effects by estimating equation (1) by groups. The groups we will create are:

- Self-employed

- Immigrant background; native, immigrant, foreign non-resident worker.
- Age-groups
- Gender
- Marital status
- Risk score

We will also use machine learning techniques to automate the search for heterogenous treatment effects. We will use all the control variables for this. There are many different types of machine learning algorithms and we have decided to use random causal forests (R package grf, Wager and Athey, 2017). As this field is moving rapidly, however, it is possible that there will be other techniques that are relevant for us once we start analyzing the data.

It will also be interesting to examine which groups that are affected in their SDR by the treatment. We will also investigate if it seems to be the case that the effects of the different treatments run via SDR.

8 Power

We are testing two main hypotheses with the 12,495 individuals (see 2). In testing the b and c coefficients, we can think of them as tests with about 8,300 individuals. At the conventional level of significance of 0.05 and a power of 0.8, our sample size would allow for a minimum detectable effect of 0.06 standard deviations. We will also adjust the p-values for the fact that we are testing the impact on two outcomes. We follow the recommendations of Fink et al. [2014] and use a method developed by Benjamini and Hochberg [1995] and Benjamini et al. [2001] to minimize the false non-discovery rate. The main advantage of the method is that it is limiting the risk of false discoveries while only adjusting the critical values based on other true hypotheses. The false discovery rate method developed by Benjamini and Hochberg [1995] implies that the m $p-$values of the i hypotheses are ordered from low to high and that the critical value of the $p-$value is then $p(i) = a * i/m$. To illustrate, with two hypotheses and a significance level (a) of 0.05, the critical $p-$value would be 0.025 for the one with the lowest p-value (0.05* 1/2, which is the same as a Bonferroni correction. For the second hypothesis, the critical $p-$value is 0.05 (0.05*2/2). The minimum detectable effect for our variable with the lowest p-value after accounting for multiple hypothesis testing (p=0.025) is 0.07 standard deviations. We conclude that our experiment is very well powered.

9 Archive and data disclosure

The pre-analysis plan is archived before the data from the survey is received and before any analyses have been made on the compliance effects data for the year 2018 (see attached letters in the Appendix). We archive it at the registry for randomized controlled trials in economics held by The American Economic Association: https://www.socialscienceregistry.org/ on October 4 2019. We expect to get the data in by October 7 and we will then start analyzing data.

References

Arun Advani, William Elming, Jonathan Shaw, et al. The dynamic effects of tax audits. Technical report, Institute for Fiscal Studies, 2017.

Alexandre Belloni, Victor Chernozhukov, Lie Wang, et al. Pivotal estimation via square-root lasso in nonparametric regression. *The Annals of Statistics*, 42(2):757–788, 2014.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57 (1):289–300, 1995.

Yoav Benjamini, Daniel Yekutieli, et al. The control of the false discovery rate in multiple testing under dependency. *The annals of statistics*, 29(4):1165–1188, 2001.

Jason DeBacker, Bradley T Heim, Anh Tran, and Alexander Yuskavage. Once bitten, twice shy? the lasting impact of irs audits on individual tax reporting. *Journal of Financial Economics*, 117(1): 122–138, 2015.

Günther Fink, Margaret McConnell, and Sebastian Vollmer. Testing for heterogeneous treatment effects in experimental data: false discovery risks and correction procedures. *Journal of Development Effectiveness*, 6(1):44–57, 2014.

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

Ronald H Hodge, Alan H Plumley, Kyle Richison, Getaneh Yismaw, Nicole Misek, Matt Olson, and H Sanith Wijesinghe. Estimating marginal revenue/cost curves for correspondence audits. In *IRS Research Bulletin, Presented at the 2015 Internal Revenue Service—Tax Policy Center Research Conference*, 2015.

Henrik Jacobsen Kleven, Martin B Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica*, 79(3):651–692, 2011.

Jeffrey R Kling, Jeffrey B Liebman, and Lawrence F Katz. Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119, 2007.

Knut Løyland, Oddbjorn Raaum, Gaute Torsvik, et al. Compliance effects of risk-based tax audits. 2019.

Luigi Mittone, Fabrizio Panebianco, and Alessandro Santoro. The bomb-crater effect of tax audits: Beyond the misperception of chance. *Journal of Economic Psychology*, 61:225–243, 2017.

The Norwegian
Tax Administration

Return address:
P.O.Box 9200 Grønland, N-0134 OSLO

Our date
24. Sept 2019

Your date

Inquiries to
Kavita Bhamra

+47 800 80 000
skatteetaten.no

Your reference

Telephone
47 97 63 6045

Org. nr:
974761076

Our reference
2018/456066

Postal address
P.O.Box 9200, Grønland
N-0134 OSLO

To whom it may concern

On exchanging survey data between the Norwegian Tax Administration and Opinion

Department of Security and Safety (DSS) in the Norwegian tax administration (NTA), is responsible for anonymization of taxpayer data collected by a private Norwegian survey institute, Opinion, on the behalf of Department of Analysis (DOA), NTA, and Oslo Fiscal Studies (OFS) at the University of Oslo. DSS is also responsible for the link key between the taxpayers social security number and a survey number constructed randomly for this particular survey. The responsibility held by DSS implies that we receive a sample of taxpayers from DOA, and conveys these for a survey by Opinion. When Opinion finish their survey, the collected data are sent back to DSS and are matched with other background variables before anonymization. The matched and anonymized data will be in the custody of DSS until January 15th 2020. After this date, the data will be made available for DOA, OFS and other interested parties.

Yours faithfully

Svein Mobakken
Director of security and safety
Security and safety department
Norwegian Tax Administration

The Norwegian
Tax Administration

Return address:
P.O.Box 9200 Grønland, N-0134 OSLO

Our date
25. Sept 2019

Your date

Inquiries to
Knut Løyland

+47 800 80 000
skatteetaten.no

Your reference

Telephone
47 92 28 0634

Org. nr:
974761076

Our reference
2018/456066

Postal address
P.O.Box 9200, Grønland
N-0134 OSLO

To whom it may concern

Self-declaration

Kotsadam, Løyland, Raaum, Torsvik and Øvrum (2019) have developed a pre-analysis plan for the analysis of a field experiment that was conducted by the Norwegian Tax Administration (NTA) in 2018 (for the tax year 2017).

The pre-analysis plan includes a section on estimating the future compliance effects of the two treatments (tax audits of taxpayers and letters to the taxpayers). This part of the analysis requires tax return data on the participants from the tax year 2018.

While Arnstein Øvrum and Knut Løyland have had access to part of the data as part of their daily work, we hereby confirm that the specific data related to this project have not yet been analyzed at the date of publishing the pre-analysis plan.

Yours faithfully

Marcus Zackrisson
Head of Department of Research and Analytics
Department of Research and Analytics
Norwegian Tax Administration