CESifo WORKING PAPERS

7381
2018

November 2018

Collusive Tax Evasion by Employers and Employees: Evidence from a Randomized Field Experiment in Norway

Marie Bjørneby, Annette Alstadsæter, Kjetil Telle

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7381*
Category 1: Public Finance

Collusive Tax Evasion by Employers and Employees: Evidence from a Randomized Field Experiment in Norway

Abstract

Third-party reporting and employers' tax withholding are powerful compliance mechanisms, as long as the employer and employee do not collude to evade. Using data from randomly assigned on-site audits among 2,462 Norwegian firms, we provide evidence of collusive tax evasion. We find that firms assigned to be audited increased their subsequent wage reporting on behalf of their employees by 18 percent relative to firms assigned to the control group. The effect is more pronounced among small firms with few employees. Our results document the limitations of third-party reporting, but also that these limitations can be counteracted by relatively inexpensive on-site audits.

JEL-Codes: E260, H260, H320.

Keywords: collaborative tax evasion, collusive tax evasion, random audits, undeclared work, third-party reporting.

*Marie Bjørneby**
School of Economics and Business
Norwegian University of Life Sciences
Norway - 1432 Ås
marie.bjorneby@nmbu.no

Annette Alstadsæter
School of Economics and Business
Norwegian University of Life Sciences
Norway - 1432 Ås
annette.alstadsater@nmbu.no

Kjetil Telle
Research Department
Statistics Norway
Oslo / Norway
kjetil.telle@ssb.no

*corresponding author

October 2018
We thank the Norwegian Tax Administration for the cooperation and access to their data. We would also like to thank Brita Bye, Edwin Leuven, Simon Quinn, and numerous seminar and conference participants for helpful comments and suggestions. Marie Bjørneby would like to thank Statistics Norway for their hospitality during the work on this paper. Financial support from The Research Council of Norway and The Nordic Tax Research Council is acknowledged.

1. Introduction

Third-party reporting and tax withholding are vital parts of the enforcement strategies in modern tax systems. Employers withholding and remitting taxes on behalf of their employees are assumed to effectively restrain under-reporting of wage income, and recent empirical studies convincingly document the power of third-party reporting in combating tax evasion (e.g. Kleven et al. 2011). However, this only works if the employer and employee do not collude to evade taxes (Yaniv 1992). This can be particularly relevant in smaller firms with few employees, where incentives and opportunities for collusive tax evasion can be strong (Kleven et al. 2016).

To our knowledge, this is the first study to use random audits to provide evidence on collusive tax evasion. Using data from on-site audits in Norwegian firms with few employees, we find that firms randomly assigned to be audited both reported more employees and higher wages relative to firms assigned to the control group. Thus, third party-reporting is no guarantee for accurate reporting as long as involved parties can coordinate underreporting of the tax base. This is important to consider for tax administrations when allocating their enforcement resources.

Analyses of prior random audit programs find close to zero effects of audits on third-party reported income, suggesting that third-party reporting prohibits evasion (Kleven et al. 2011; DeBacker et al. 2015a; Advani et al. 2017). However, results from such studies provide lower bounds of tax evasion, as they rely on *detected* tax evasion. Kleven et al. (2011) underline that the desk-audits used in their study are largely incapable of detecting evasion of labour income from the informal economy. Underreporting of wages is difficult to detect if the employer and the employee align their reports to jointly underreport income. As discussed by Kleven et al. (2016), while collusive tax evasion can be hard to sustain in firms with many employees and accurate business records, evasion can be more easily coordinated in small firms in the fringes of the formal economy where the interests of the employers and employees largely overlap.

By collusive tax evasion we mean that the employer and employee coordinate, possibly tacitly or even unintentionally or unwillingly, to not report the (full) tax liability. The collusion may even be perceived as involuntary by one party. For example, for unskilled or illicit workers, the only perceived alternative to accepting undeclared wages could be to exit the labour market and the country. Nevertheless, in our setting where both the employer and employee are legally required to report the income accurately and the reports are matched upfront through pre-filled tax returns, successful evasion would require some sort of coordinated misreporting.[5] How the employer and employee split the gain will depend on the demand and supply elasticities, but also on their attitude towards risk and how the legal responsibility for tax payment is divided between them (Sandmo 2012), as well as their relative bargaining powers.[6]

The main enforcement strategy to detect undeclared work is unannounced on-site audits that determines the identity of all persons present. However, if undeclared employees are discovered, firms could claim that the person was just hired. In a series of inspections by the Danish Customs and Tax in 2004, one third of unregistered employees claimed that it was their first working day in the firm (Kolm and Bo Nielsen 2008). To avoid this, the Norwegian government introduced a new monitoring rule to combat undeclared work from 2014, requiring firms to maintain a real-time staff register of every person present at the workplace at any time. These registers ought to be available at the site for unannounced on-site audits. We utilize this new requirement to document collusive tax evasion.

In order to analyse the effects of the new 2014 staff register audits, we cooperated with the Norwegian Tax Administration to design and implement a field experiment where 2,462 firms required to keep a staff register were randomly assigned to an audit group and a non-audit group (Thorsager and Melsom 2017). The audits were unannounced, on-site, and directly targeted at detecting undeclared work by examining whether all employees present at the time of audit were registered in the staff register. These on-site audits, although relatively non-extensive, are well suited to detect unreported personnel and associated labour income from the informal economy, and thus to have a stronger deterrence effect on wage underreporting than the desk-audits used in many previous studies.

By linking the randomized audit data to administrative micro data we can analyse the causal effects of these on-site audits on firms' subsequent wage reporting. If audited firms perceive that the probability of detection of wage-underreporting is increased, we would expect an increase in wage reporting following an audit. We find that firms assigned to be audited on average increased their wage reporting on behalf of their employees by 18 percent and the number of reported employees by 22 percent, relative to firms assigned to the control group.

---

[5] In the Norwegian setting, where employees' tax returns are prefilled by the tax authorities based on reports from the employer, collusive tax evasion may only require that the employee remain ignorant about the correct tax base and thus do not correct it. Such passive behaviour, is, nevertheless, illicit.

[6] In a different setting, Nygård et al. (2016) study how consumers and suppliers of goods and services split the gain from joint tax evasion.

Our findings suggest that on-site audits, even if they are non-extensive and inexpensive, can be a necessary supplement to desk-audits, in order to increase compliance especially for small firms with few employees.

Section 2 reviews the relevant literature, while Section 3 presents the institutional setting, discusses the incentives for participating in collusive tax evasion, and the design of the randomized audit experiment. Section 4 describes the data and Section 5 the empirical strategy. Our results are presented in Section 6 and Section 7 concludes.

2. Related international literature

2.1 Tax evasion with third-party reporting and withholding

The standard theory on tax evasion builds on the deterrence-model by Allingham and Sandmo (1972). The model predicts that tax evasion depends on taxpayers' perceived probability of detection and the punishment if detected compared to the gain of reduced tax payment.

Extending this model to incorporate third-party reporting and withholding taxes, one could argue that the probability of detection should be (close to) 100 percent. However, when the employer and employee collude to evade, third-party reporting will not eliminate evasion. Evasion of taxes on wage could include failing to register the firm, failing to register some or all employees or underreporting wage payments for registered employees (e.g. declare the minimum wage and pay the rest in cash). Collusive tax evasion may be extremely hard to detect in audits and represents a potential threat to the success of third-party reporting.

Even if third-party reporting and withholding will not eliminate evasion, it will still increase the cost of underreporting because it requires an (at least tacit) agreement between the two parties involved. Hence, if an audit is not able to detect evasion, there will be a risk of detection connected to whistleblowing, which is likely to increase with the number of individuals involved. Hence, collusive tax evasion is easier to maintain in small firms with few employees (Kleven et al. 2016). Close relationships between the parties involved, as in closely held firms and family firms, would make collusion even less risky. In a recent paper, Barth and Ognedal (2017) argue that not only the risk of whistleblowing, but also the cost to the firm of whistleblowing increases with firm size.

Undeclared work is likely also linked to tax evasion in other dimensions, which can also be more prevalent in some sectors than in others. Especially, some firms can easily hide cash income and cash expenses from the tax administration by keeping two sets of financial records – one fraudulent official record and one true for personal records (unavailable to the government).[7] Such a "parallel cash economy" is easier to maintain for firms involved in a large number of small-value sale transactions to individual customer (rather than business-to-business sales) who more often pay in cash and less often demand a receipt (Slemrod et al. 2017; Kleven et al. 2016; Pomeranz 2015; Santoro 2017).

Collusive tax evasion could potentially have different features and respond differently to enforcement than individual tax evasion. Abraham et al. (2017) argue that because collusive evasion involves social interaction, social norms have a stronger negative effect on collusive evasion than on individual evasion. They confirm these predictions in a laboratory experiment. Furthermore, higher penalties could sustain collusive tax evasion because it increases the cost of breaking the collusion (Boadway et al. 2002).

2.2 Effects of audit on subsequent compliance

In addition to raising revenue directly by detecting tax evasion, audits could potentially also affect future compliance of audited taxpayers. It is not given *a priori* what the effects of audits on future compliance are. Being audited could affect the taxpayers' perceived probability of detection in the future in (at least) two ways. First, being audited could alter the perceived probability of being audited again in the future. And, second, being audited could also affect the perceived "audit productivity" – the probability of detection if audited.

Considering the perceived probability of future audits, there are contrasting theories on how taxpayers update their audit probability following an audit. The "target effect" predicts that audited taxpayers revise their perceived audit probability upwards after an audit because they believe that being audited indicates that tax administrations consider them to be potential evaders and that they will be more closely monitored in the future (Hashimzade et al. 2014). Several randomized field experiments have found that tax audits have a positive effect on future compliance by individual taxpayers (e.g. Kleven et al. 2011; DeBacker et al. 2015a; Advani et al. 2017). On the other hand, the "bomb crater effect" predicts that taxpayers revise their perceived audit probability downwards after an audit because they believe that it is unlikely to be audited again in the near future (Mittone 2006). DeBacker et al. (2015b) and Di Porto (2011) both find that corporations increase their tax aggressiveness after audit.

[7] To combat this form of evasion, many countries have introduced certified cash registers where all sales are registered in real time.

Considering the audit productivity, a taxpayer will update her perceived audit productivity based on the experience of previous audits. If an audit failed to detect evasion, or if the penalty was milder than expected, the taxpayer might reduce compliance and pursue more of what she now considers to be less risky evasion. Gemmell and Ratto (2012) find support for such a disparate effect that depends on the tax administration's success in detecting noncompliance. They find that audited taxpayers found to be compliant reduce their subsequent compliance, while taxpayers found to be noncompliant increase their compliance.

2.3 Randomized audits

Studies using random audits have the advantage of avoiding the identification problems and selection bias associated with studies using observational data. Generally speaking, such random audit programs have focused on desk audits towards individual taxpayers (wage earners and self-employed), see e.g. Kleven et al. (2011); DeBacker et al. (2015a); Advani et al. (2017). However, recent studies have found that on-site personal visits by a civil servant tend to have a larger effect on compliance than both desk-audits (D'Agosto et al. 2018) and receiving a letter (Telle 2013; Dörrenberg and Schmitz 2017; Boning et al. 2018).

Boning et al. (2018) rely on a field experiment and find that on-site personal visits by a civil servant have a substantial immediate effect on tax remittance by US firms that seemed to be falling behind on their tax deposits of employment taxes. The US Internal Revenue Service does not match tax returns against third-party reports until long after the filing deadline. In such a setting, unilateral noncompliance by one party without the collusion of the other, could be more prevailing (i.e. employers‘ theft of taxes withheld). This is very different in Norway, where personal tax returns are pre-filled by the tax authority with income reported by third parties (employers, banks, etc.).[8]

[8] There are other papers that examine employers‘ evasion of employment taxes, with or without collusion of employees, using operational (i.e. non-randomized) audits (Di Porto 2011) or changes in tax rates (Madzharova 2013), pension rules (Kumler et al. 2013) and minimum wage regulations (Tonin 2011) as natural experiments. Besim and Jenkins (2005) find that private employees underreport incomes to the same degree as do self-employed (using the food expenditures approach pioneered by Pissarides and Weber (1989), but with civil servants as benchmark).

3. Institutional setting

3.1 Third-party reporting and remittance of wages

In Norway, as well as in the other Nordic countries, personal income tax return preparation has been fully automated for the vast majority of taxpayers. In 2013, 63 percent of the returns were fully pre-filled and 91 percent of personal income tax returns were filed electronically (OECD 2015). These pre-filled tax returns are based on extensive third-party reporting.

Employers are required to withhold and remit income tax and social security contributions from employees' pay checks.[9] The employer is required to document this remittance by sending a receipt (wage slip) to the employee, stating the before-tax wage and the amount of the taxes withheld for each wage payment and at year-end.

Employers must also report details of individual employees' wages directly to the Norwegian Tax Administration on an annual basis. The Tax administration uses this information to pre-fill the tax returns made available to taxpayers for validation. Despite being pre-filled the taxpayers are still legally responsible to examine all the information and report inaccuracies to the Tax administration. If no changes are reported by a given date, the tax return is considered filed ("deemed acceptance").

The pre-filled returns (upfront matching) provides an efficient tool to screen individuals' tax returns, where any form of unilateral noncompliance by employers or employees is likely to be detected with minimal enforcement resources required.

If an employer fails to actually remit to the Tax administration the amount of taxes he has withheld from the employees' wages, the wage slip serves as a documentation of remittance on behalf of the employee and the employer is responsible for paying the taxes owed (and key personnel within the firm can be held financially and criminally liable if the firm has gone out of business). If the employee cannot provide such a wage slip to document withheld taxes, the employee is liable for the missing tax payments. The employee thus has a strong incentive to check that the wage he receives is what the employer reports to the Tax administration. Hence, we argue that unilateral noncompliance by one party without the collusion of the other (i.e. employer's theft of taxes withheld), is not likely to be prevailing in our setting. This is different from the situation in many other countries, where tax returns are not automatically matched

[9] The required amounts of withholdings are provided in individual tax cards based on the taxpayers' historical income and tax assessments. The taxes are withheld from the employee's pay check and deposited in a separate bank account until it is remitted to the Tax administration every second month.

against third-party reports (or are only matched after the filing) and where discrepancies will only be detected through (desk) audits.

3.2 Incentives for participating in collusive tax evasion

Assume, first, that the firm follows other regulations and reports actual turnover and pays corporate taxes accordingly. Then, underreporting wages will reduce the consolidated tax burden of the two parties if the total marginal tax rate on wages exceeds the tax rate on corporate profit.

By under-reporting wages, the employee escapes income taxes (27 percent in 2014) and social security contributions (8.2 percent) and can agree to a wage below his before-tax market wage. The employer foregoes the tax deduction for the undeclared wage costs resulting in overpayment of profit taxes (27 percent), but benefits from reduced pre-tax wage costs and reduced social security contributions (14.1 percent). Overall, the maximum marginal wage tax of 47.2 percent is well above the corporate profit tax of 27 percent, which implies that there are incentives on the margin to shift income from the former to the latter tax base.

Incentives to under-report wages are even stronger if the collusion allows the employer to claim deduction for unreported wages or suppress sales (to evade VAT and profit taxes). Collusion could also involve tax evasion in other dimensions. If a firm underreports (cash) income and uses that income to pay undeclared wages to employees, both personal income taxes and payroll taxes as well as corporate income tax and VAT are evaded. The firm and employees have to balance the benefits of reduced administrative costs (compliance costs and other administrative costs imposed by government regulations) against the costs of reduced legal rights and social security rights and reduced access to credit and markets.[10]

Moreover, if caught under-reporting, the employee faces penalty taxes and potential criminal prosecution. The regular penalty tax is 30 percent of taxes evaded (in 2014), but it is reduced to 10 percent if the information is correctly reported by a third-party. An additional penalty tax of 15 or 30 percent applies if the under-reporting is done with clear intent ("wilful failure"). Tax fraud is also punishable with fines or prison for up to 6 years. A third-party that misreport or fail to withhold and remit taxes, can be punished with a fine or prison. If the third-party is a firm, the personal responsibility for misreporting could be on the board, the director or the person actually performing payment and withholding.

[10] Reduced social security contributions will not necessarily reduce the social security-rights correspondingly (i.e. due to ceilings on social security benefits such as pension and unemployment benefits).

3.3 Experimental design

In 2014, the Norwegian government introduced a new monitoring rule to combat undeclared work. Firms in certain service sectors with a high degree of private customers and possibility for cash turnover - food service, hairdressers, beauticians, car repair and car care-businesses - are required to maintain a staff register, recording every person present at the workplace at any time (Thorsager and Melsom 2017). These sectors were selected because they were considered by the Tax authority to have relatively high opportunities to engage in tax evasion. Also, these firms have established places of business, making it possible to inspect the people on-site at work in an unannounced audit.

The staff register should contain the name and ID of everyone working at site and when they started and ended their work each day. The register ought to be available at site for unannounced audits. This provides an opportunity to monitor employment and identify businesses that violate the regulations, both for the Norwegian Tax Administration, the Norwegian Labour and Welfare Administration and the Norwegian Labour Inspection Authority.

In cooperation with the authors, the Tax Administration in Oslo designed and implemented a field experiment with firms randomly assigned to be audited. The audits were on-site and directly targeted at detecting undeclared work. The auditors examined the firms' staff registers and compared it to the people present at the workplace at the time of audit.

The experiment is based on a base population of 2,462 firms. This is the overall population of firms in the Oslo region which were assumed required to keep a staff register in January 2014. When designing such experiments, there is a trade-off between the tax administrations' need for higher audit probability among firms they consider "usual suspects" who are highly likely to evade, and the methodological need for randomization in order to ensure identification and credible estimates. A pragmatic solution to this is stratification (Telle 2013), and the Tax Administration thus divided the firms into 12 strata, according to branch, size and a subjective risk assessment. This enabled the Tax Administration to follow its risk-based audit policy, i.e. to place firms that they believed were most (least) likely to violate into one stratum and audit a higher (lower) proportion of firms in this stratum (see Table 1). In the literature, this experimental design is often referred to as a "blocked" or "stratified" randomized experiment; see e.g. Athey and Imbens (2017).

In each stratum, a varying proportion of firms were randomly assigned to the treatment group for unannounced audits of staff registers. In total 923 firms were drawn to be audited. The remaining 1,539 firms were assigned to the control group and were not supposed to be audited. The audits were carried out in 2014 and the first half of 2015. Audits were unannounced and the taxpayers were not told that the audits were part of an experiment.

If the auditors discovered irregularities in the staff register, the firm was fined. Also, more than half of the violating firms received a new on-site audit. The fine for failing to maintain the register was about 1,000 € (NOK 8,600) for the first offence and 2,000 € for further offences within 12 months. In addition, a fine of 200 € was imposed for each person not included in the register at the time of audit. Since the vast majority of these firms are very small (see Section 4.3. below), the size of these fines are not negligible.

The main deterrence effect of these audits may not necessarily be the risk of being fined for failing to maintain the staff register, but that the Tax Administration would pursue irregularities and uncover undeclared work and turnover through more comprehensive audits. While we did not access information on such actions undertaken by the Tax Administration, our results should be interpreted in light of possible follow-up of suspicious firms in line with standard procedure of the Tax Administration.

Given that these audits have a deterrence effect, it is still not given what the observed response would be. The most obvious expectation would be that evading firms will *increase* their reporting if they perceive that the probability of detection is increased. But being audited could also induce evading firms to close down their activity and workers to move to other formal or informal firms or to exit the labour market, which would appear as *reduced* reporting in our data. Hence, the response we observe could include both positive compliance responses and, to some extent, negative real responses, which will give a downward bias in our estimates.

Another possible real response is that our audits detected workers that were paid below the minimum wages, and hence that the effect we see measures a real increase in wages.[11] However, the fact that the reported number of employees also increased (and, relatively more than the wage increase) indicates that this is not the main effect. Audits could also induce firms to hire a new employee to deal with the Tax Administration and to manage the staff register, but this seems unlikely since it is simple to set up a staff register (it is sufficient that the employee writes his/her name on a paper – or in a simple electronic app - when present).

[11] In Norway, there is no general minimum wage, but a general application of collective agreements concerning pay and working conditions has been introduced in certain sectors (i.e. restaurants).

To separate real responses from evasion responses is a general challenge when analysing behavioural responses to taxation (Slemrod 2001). In a more general setting, this could also affect labour market outcomes for others, by affecting the competition between evaders and compliers, wages and prices, and the industry structure etc. (see e.g. Kolm and Bo Nielsen 2008; Kolm and Larsen 2018).

Another reason why our effect estimates on the treated could be downward biased, is if audited firms revealed information about the audits to non-audited firms who then reduced evasion. Given that our experiment only covers firms in specific sectors in the Oslo region, such spillover effects through information sharing from the treatment group to the control group cannot be ruled out. The general public awareness and media coverage following the introduction of the staff register requirement could also potentially have led to reduced tax evasion by firms both in the treatment and the control group.

4. Data description

4.1 Data from the experiment

We have a list of all 2,462 firms that were part of the experiment, with information on stratum and whether the firm was randomly assigned to the treatment group (923 firms) or the control group (1,539 firms). Furthermore, we have all the reports from the on-site staff register-audits of these firms in 2014 and the first half of 2015, with information about the audits, such as start and end date, the outcome of the audits ("accepted" or "fined", and the size of the fine if applicable) and any annotations made by the auditor.

The audits revealed extensive violation with the staff register requirement. In the first audit, 25 percent of the firms received a fine, 28 percent received a warning but no fine and 47 percent were accepted. Around half of all the firms received a second audit. In the second audit, 30 percent of the firms that received a fine in the first audit were fined also the second time. Interestingly, among the firms found in compliance in the first audit, 20 percent were fined in the second audit. Since the second audit is not randomly assigned, we cannot infer any causal effect from these findings. We do not know whether these firms did actually comply in the first audit, or if (some of) these firms were in fact in violation also in the first audit without being detected. These findings may also suggest that the auditors follow up with a new audit in firms for which they (correctly) suspect (but cannot satisfactorily document) violation. Altogether, one third of the audited firms were found to violate the staff-register regulations (one or more times). This violation may not have been deliberate; it could be due to lack of information about the new regulation. And more importantly, violating the staff-register regulations is not necessarily linked to undeclared work and tax evasion. Our analysis does not depend on the outcome of the audits but on how the firms alter their reporting after being audited.

The audit reports reveal that a number of firms randomized for audit did in fact not receive one. For 39 percent of the firms in the treatment group, we do not have any audit report, indicating that the firm was not audited. For additionally 23 percent of the firms, we have audit reports but the outcome of the audit ("accepted" or "fined") is lacking, suggesting either that the audit for some reason was cancelled or that the auditor discovered minor irregularities in the staff register, but chose to give a warning and not to levy a fine. In these cases, with missing outcome of the audit, we used annotations made by the auditors to distinguish between assumed audited and not audited firms. In total, we find that only around half of the firms randomized to be audited were actually audited (see Table 1).

There could be several reasons why a number of the randomly assigned firms were not audited. Clearly, the firms could not choose not to be audited and the auditors were instructed to audit all firms in the treatment group. The sectors required to maintain the staff register - food service, hairdressers, beauticians, car repair and car care-businesses - are dominated by small firms that tend to be short-lived. Firms that ceased to exist, moved out of the jurisdiction of the Oslo Tax administration or changed their business activity, could not be audited. Some firms may also have been mislabelled in the registry from which they were drawn due to e.g. incorrect sector classifications.

Several biases can occur from this. What firms in the treatment group that are in fact audited may not be random, contributing to upward bias in effect estimates if they are more prone to violations and downward bias if they are less prone than the audited firms. More subtly, while firms in the treatment group may not be audited because they are carefully investigated and found not required to have a staff register, similar firms in the control group are not identified. Thus, excluding the non-required (and thus non-audited) firms in the treatment group from the sample would introduce a non-transparent selection with accompanying bias, because we cannot identify and exclude similar firms in the control group.

An additional complication is that firms randomly assigned to the control group were in fact audited. This is a common challenge with real field experiments where tax authorities have to balance the benefits of a randomized experiment and the drawback of leaving tips of suspected evasions (received after randomization) go unattended. In our main analytic sample (described in Section 4.3), 47 of the 1,354 firms in the control group were audited. This means that the audit frequency of the control group (3 percent) is not negligible, though it is small compared to the 48 percent audit frequency in the treatment group (see Table 1).

If we are after the effect of the intervention as a whole (intention to treat), such auditing of firms in the control group that is part of the counterfactual operation of the Tax Administration is not biasing the estimate. But if one attempts to estimate the effect of the intervention on the firms in the treatment group (treatment effect on the treated), such contamination of the control group contributes to a downward bias, since firms in the control group are actually audited. Unfortunately, such contamination is also likely to reduce the correlation between being assigned an audit and actually being audited, which will hamper the precision of estimates of local average treatment effects (LATE), as discussed in Section 6.3.

4.2 Administrative data

To analyse the effect of audit on firms' tax compliance, we merged the data from the experiment with administrative data, relying on unique firm identifiers available in all registers in Norway. We accessed yearly data for 2009-2014 on wage payments by all Norwegian employers to all employees on employment level from The Register for Pay and Tax Deducted ("Lønns- og trekkoppgaveregisteret"). This information is reported to the Tax Administration by employers and used for making pre-entries in employees' tax returns. Variables include firm and employees' IDs, wage payments and other taxable benefits from the firm during the year. Also, we add firm characteristics variables from The Register for Legal Entities ("Enhetsregisteret"): organizational form, the year the firm was established and, if applicable, deleted.

From these data we construct three outcome variables on firm-year level: (1) wages, (2) number of employees (all individuals reported to receive wage from the firm), and (3) a dummy variable indicating whether or not the firm reported any wage. If the audits had a deterrence effect on collusive tax evasion, we expect wage reporting to increase. If evasion also comprised the use of undeclared workers, we anticipated an increase in the number of reported employees. Furthermore, if some firms failed to report wages, we expected the fraction of firms reporting to increase.

Half of the firms did not report anything on our outcome variables in 2014. The choice not to report might be an effect of the audit, even if there is no statistically significant difference in the propensity to report in the treatment and control group. These firms are thus included in the analysis with the outcome variable set to zero. This implies that we measure the total effect of audits, including any potential effects on the extensive margin through firms exiting or starting to report. The audits were carried out during all of 2014 and the first half of 2015. The annual deadline for employers to report wages for the fiscal year t is the end of January in t+1, one month after the end of the fiscal year t, but reports can be adjusted until March 1st in t+1. Audits carried out until March 1st in year t+1 may affect the reporting for year t, given that firms have the opportunity to modify their reports immediately. Firms also report wages and remit withholding taxes throughout the year. However, this is reported on the firm level only, and we assume that firms can easily increase their final tax reporting on the employee level at year end if they are deterred after being audited. Thus, reports for the fiscal year 2013 could be affected by the audits carried out in the beginning of 2014 (January-March).

As pre-reform information, we therefore have to rely on the reports for the fiscal year 2012 to be sure that we have information not affected by the audits. We rely on reports for the fiscal year 2014 to measure effects of the audits on the outcomes, though this may result in some downward bias in effect estimates as 7 percent of the audited firms in the treatment group did not received the audit by March 2015.

4.3 Sample restrictions and descriptive statistics

The population of 2,462 firms was defined by the Tax Administration as the overall population of firms in the Oslo region required to keep a real-time staff register. This was done by extracting all firms in the relevant sectors (according to NACE-codes) and region from The Register for Legal Entities in January 2014.

This is not a completely accurate identification of firms required to keep the real-time staff register. First, firms could have activities within different sectors, and the requirement to keep a staff register depends on the real activity within each sub-unit (special rules apply to mixed activity). Furthermore, this does not capture unregistered firms that are unobserved by official authorities (informal or underground economy). To cover more of the relevant firms, the Tax Administration also added businesses that had license to serve alcohol. Second, there are many passive firms in The Register for Legal Entities, and there is a trade-off between including or not including firms with no reporting. If these firms have closed down their activity, audit-resources would be wasted. However, if these firms are doing businesses without reporting it, these are exactly the firms the Tax Administration would like to audit. Whether such firms were or were not included in the population was based on judgement by the Tax Administration before the random assignment.

In our analysis, we include non-reporting firms. But we exclude 42 firms that were deleted from The Register for Legal Entities before the experimental audits started. The Tax Administration did not attempt to audit these firms, and the close-down of the firm could thus not possibly be an effect of being assigned to be audited. Furthermore, we exclude 65 firms that occurred several times in the Tax Administration's list used for random assignment, and hence had a higher probability of being assigned to the treatment group than the rest of the firms.[12]

We are then left with a sample of 2,355 firms. As shown in Table 2, the majority of these firms are very small. The median firm reported only 1 employee in the year before the audits started (i.e. 2012). Note that almost half of the firms did not report any wage or employees. However, this sample also contains some comparably very large firms.

To focus on firms where collusive tax evasion is likely to take place and to reduce the impact from some outlying very large firms, we excluded the 10 percent largest firms (i.e. firms reporting > 26 employees in our baseline year, 2012) from our analytic sample. The remaining sample, which we will refer to as our *main analytic sample*, will be used in the analyses (unless otherwise noted) and contains 2,117 firms, as shown in Table 2.[13]

As noted in Section 3.3, randomization took place within 12 strata; see Table 1. There is substantial heterogeneity across strata, as the number of firms in each stratum varies from 9 to 551, and the mean number of employees varies from 0.18 to 12.6. Moreover, the share of firms within a stratum assigned to treatment varies between 4 percent and 71 percent, and the share of firms actually audited among the firms assigned an audit varies between 12 and 100 percent. This heterogeneity suggests that the effect of being assigned to audit could also vary substantially across strata, in which case it will be important for the interpretation of our findings how results from each stratum are weighted, as we will discuss in Section 5.

[12] The reason for such duplicate firm IDs is that one firm may have several geographically different points of sale, and each such point of sale may then be included with its sub-ID in the population, even if they belong to the same legal entity with identical firm ID.

[13] We also conduct the analysis for the whole sample and for subsamples with fewer employees in 2012, and the estimates are reported in Appendix Tables A.1 – A.3.

Table 1 Descriptive statistics by strata

Strata	Number of firms	Fraction of firms reporting wages	Mean wage (mill NOK)	Mean number of employees	Fraction of firms assigned to treatment group	Fraction of firms in the treatment group that were actually audited	Fraction of firms in the control group that were actually audited
1	50	0.56	0.51	6.76	48 %	25 %	8 %
2	92	0.24	0.25	1.82	70 %	27 %	21 %
3	250	0.36	0.56	2.21	19 %	15 %	1 %
4	271	0.82	0.79	4.46	71 %	47 %	3 %
5	53	0.17	0.06	0.55	64 %	12 %	0 %
6	482	0.06	0.02	0.18	4 %	15 %	1 %
7	551	0.74	0.65	6.63	55 %	65 %	7 %
8	52	0.62	0.33	4.48	29 %	73 %	3 %
9	198	0.33	0.11	2.08	5 %	67 %	2 %
10	9	0.67	0.90	12.56	44 %	100 %	0 %
11	39	0.62	0.70	6.03	8 %	33 %	0 %
12	70	0.33	0.29	3.77	66 %	50 %	29 %
Total	2,117	0.45	0.41	3.45	36 %	48 %	3 %

Note: Number of firms, fraction of firms reporting positive wages, average reported wage and reported number of employees in 2012, as well as fraction of firms assigned to treatment group and fraction of firms that actually end up being audited in treatment and control group, by strata. Main analytic sample cf. Section 4.3.

Table 2 Descriptive statistics

	mean	sd	p25	p50	p75	p95	N
Panel A: All firms							
Number of employees	15.69	113.38	0.00	1.00	8.00	47.00	2,355
Wage mill NOK	2.66	31.19	0.00	0.01	0.88	5.04	2,355
Fraction of firms reporting wages	0.51	0.50	0.00	1.00	1.00	1.00	2,355
Age of firm	9.27	9.27	2.09	6.19	14.23	26.35	2,355
Panel B: Main analytic sample (excluding 10 percent largest firms in 2012)							
Number of employees	3.45	5.77	0.00	0.00	5.00	17.00	2,117
Wage mill NOK	0.41	0.82	0.00	0.00	0.48	2.02	2,117
Fraction of firms reporting wages	0.45	0.50	0.00	0.00	1.00	1.00	2,117
Age of firm	8.89	9.04	1.96	5.69	13.76	25.44	2,117

Note: Age of firm is defined as years from registration to 1.1.2014. The other numbers are for 2012. For sample definitions, see Section 4.3.

5. Empirical strategy

Because firms were randomly assigned, audited and non-audited firms should be statistically indistinguishable pre-treatment (2012), and the effect of audit can then be estimated by comparing outcomes across the two groups post-audit (2014). We start by estimating the average causal effect of the audits by comparing firms' outcomes in the treatment group relative to the control group post-treatment (2014).

By linking the audit-data to data on outcomes reported by the same firms from both before and after the intervention, we also compare *changes* from 2012 to 2014 in the outcomes across the treatment and control group (first-difference). This approach could help to improve the precision of our effect estimates and to correct for any random imbalance (in levels) pre-treatment. This difference-in-differences design does not require the two groups to be equal before treatment, but rests on the somewhat weaker assumption that, absent treatment, the *change* in the outcome variable in the control group is a good estimate for the counterfactual change in the treatment group. Similarly, we also show results from a model where we control for the lagged (2012) value of the outcome variable (McKenzie 2012).

Only about half of the firms assigned to the treatment group were actually audited and a number of firms were audited even though they were assigned to the control group. By including all firms in our treatment and control group, whether or not they were actually audited, we estimate the effect of being assigned an audit, i.e. the intention to treat effect (ITT). While this can be considered the estimate most relevant to capture the effect of the policy intervention, it is a downward biased estimate of the effect of actually being audited.

We have also estimated the local average treatment effect (LATE) for firms that were actually audited, instrumenting actually being audited with being assigned an audit. This measures a causal effect of being audited using the exogenous variation in audit probability generated from the random assignment as an instrument.

As discussed in the previous section, randomization was conducted within each of 12 strata with the proportion of treated varying substantially across strata. One common motivation for such a design is that stratification could increase precision if it is based on characteristics that are correlated with the outcome variable (Athey and Imbens 2017). In our experiment, however, the main motivation for stratification was to meet the Tax Administration's need for risk-based audit policy.

This stratification has two important implications for our analysis. The first being that treatment is not random in our overall sample, only within each stratum. Second, to the extent that the Tax Administration is actually correct in their beliefs about risk, the strata with high audit-rates will contain a substantially larger proportion of violators than other strata. Consequently, we suspect the treatment effects to also be heterogeneous across strata. To get a consistent estimate of the average treatment effect in the population, we thus need to weight the treatment effect within each stratum by each stratum's share of the population.

In principle, we have 12 separate randomized experiments, one within each stratum. By including a full set of interaction of treatment and strata dummies, i.e. allowing for different treatment effects within each stratum, we can estimate stratum-specific ITT in the following linear model:

(1) $$y_i = \sum_{j=1}^{J} \beta_j \times C_{ij} + \sum_{j=1}^{J} \tau_j \times Treat_i \times C_{ij} + u_i$$

where y_i denotes our outcome variable (the level of, or in the first difference setting, the change, in wage, number of employees and a dummy for whether the firm reports positive wages), $Treat_i$ is a dummy variable indicating whether firm *i* was randomly assigned an audit, C_{ij} are strata dummies and u_i is an error term with conditional expectation zero. Then, τ_j is an unbiased and consistent estimator for the ITT in stratum j.

However, we are interested in the ITT for the whole population, not the effects within each stratum (some of the strata are very small, and even if we were interested in the effect within a stratum, our experiment does not have enough power within each stratum).[14] Thus, we calculate the weighted average of the within-stratum average treatment effects, with weights being the share of firms in each stratum.[15]

Turning to the local average treatment effect (LATE) of actually being audited, our baseline IV-model can be described by the following two-equation system, instrumenting actual audit with being assigned to treatment group:

(2) $$Audited_i^j = a^j + b^j \times Treat_i + u_i^j$$ one equation for each j=1,..,J

(3) $$y_i = \sum_{j=1}^{J} \beta_j \times C_i + \sum_{j=1}^{J} \tau_j \times \widehat{Audited}_i^J + \varepsilon_i$$

where $Treat_i$ is a dummy variable indicating whether firm *i* was assigned to the treatment group, $Audited_i^j$ indicates whether the firm was actually audited, and C_{ij} are strata dummies. We estimate the parameters τ_j by performing a 2SLS with the vector of equations (2) as the first stage and equation (3) as the second stage. Similar to above this provides one LATE for each stratum, and we get the average over all strata, by weighting by the share of firms in each stratum.

While our reduced form ITT estimates can be given causal interpretation as long as the assignment was random, the IV estimates rely on three additional assumptions. First, our instrument (*Treat*) should only affect the outcome variables (*y*) through the probability of being audited (*Audited*). This exclusion restriction is likely to hold, as the random assignment is only

---

[14] Formal tests confirm that the effects are clearly statistically significantly different across strata in our data.

[15] This can be done by running OLS on Eq. (1) and weight the estimates and calculate the standard errors, or by transforming the parameter vector as described in Appendix A.0.

observed by the firm through the audit. Second, being assigned to the treatment group (control group) should increase (decrease) the probability of being audited for each firm (monotonicity assumption). It seems plausible that audited firms in the control group would also be audited if they were assigned to the treatment group, and similarly that non-audited firms in the treatment group would not be audited if they were assigned to the control group. Third, being randomly assigned an audit (i.e. being in the treatment group) is a good predictor of actually being audited. As previously noted, some of the firms in the control group were in fact audited, and half of the firms in the treatment group were not audited. Moreover, the fraction of firms in the treatment group that were not audited, and the fraction of firms in the control group that were audited, varies considerably across strata. This weakens the correlation between being assigned to treatment group and receiving an audit, particularly in some strata. As a result, we cannot estimate the LATE precisely (Angrist and Pischke 2009, p. 209), and our results may also suffer from weak instrument bias (toward OLS), including too small standard errors. We return to this in Section 6.3.

6. Results

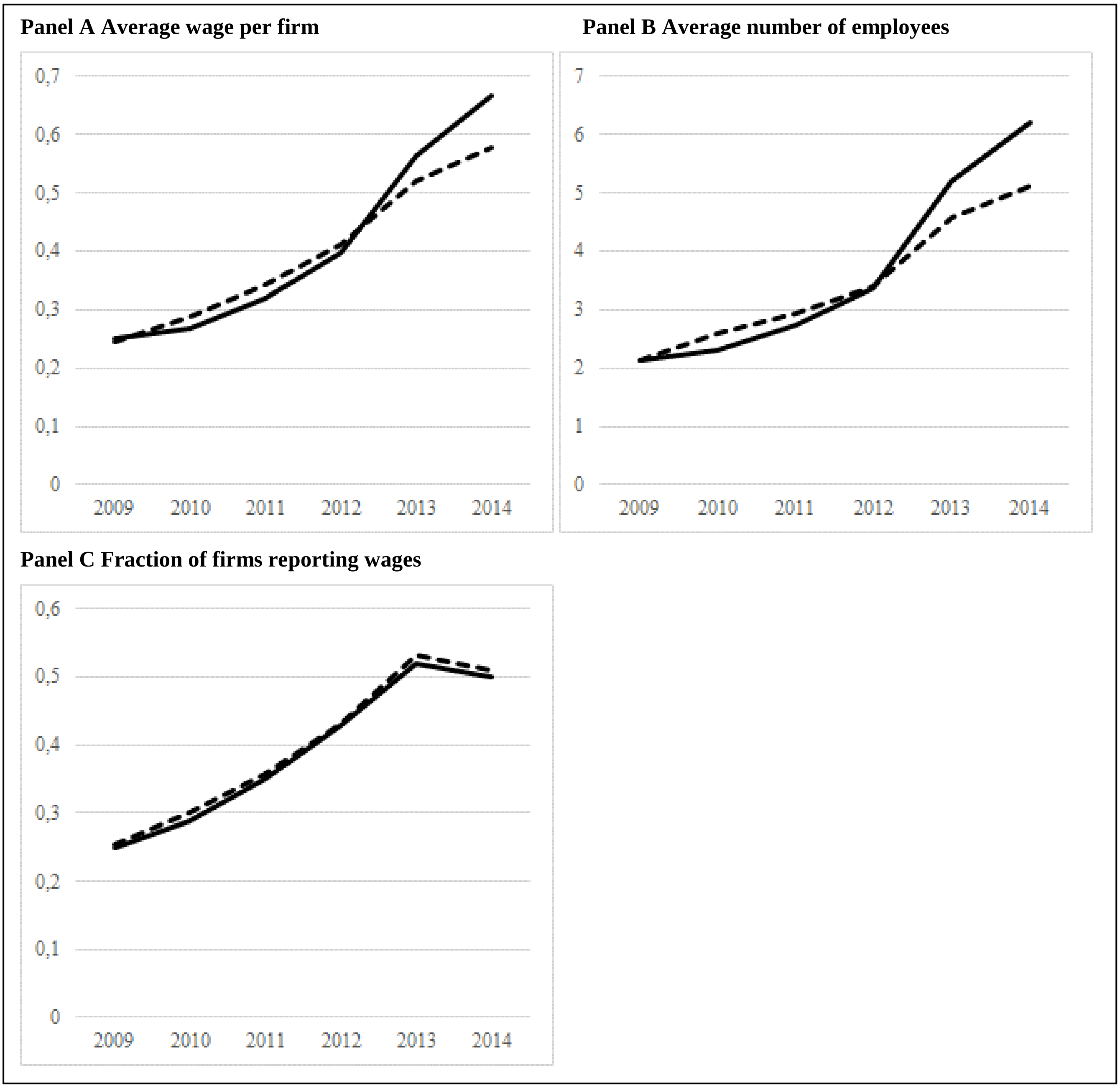
6.1 Graphical evidence

The simplest way to test the effect of being audited on subsequent reporting is to track the reporting by firms in treatment group and control group over time. Figure 1 plots means of wages, number of employees and the fraction of firms reporting wages in treatment and control groups over the 2010-2014 period.

We observe the same level of reporting up until 2012 (pre-treatment). From 2013,[16] firms in the treatment group increased reported wages (Panel A) and number of employees (Panel B) relative to firms in the control group, indicating a positive effect of audits on firms' subsequent tax reporting. However, the fraction of firms reporting any wage (Panel C) does not seem to differ between the treatment and control groups, indicating that the audits did not affect the likelihood that the firms would report any wages.

[16] Recall from Section 4.2 that reports for the fiscal year 2013 were filed in the beginning of 2014, i.e. after the experimental audits had started.

Figure 1 Average wage, number of employees and fraction of firms reporting wages in treatment group (solid line) and control group (dashed line) 2009-2014. Main analytic sample



6.2 Effects of being assigned an audit (ITT)

Our regression results, presented in Table 3, confirm that being assigned an audit had statistically significant positive effects on both reported wages and number of employees after the audit. The table shows averages for reported wages, reported number of employees and the fraction of firms reporting wages in 2012 (pre-treatment) and 2014 (post-treatment) for the treatment and control groups separately. The third column of Table 3 shows the differences in the outcome variables between the treatment and control group in 2012 and 2014, as well as the differences in the changes from 2012 to 2014.

Prior to treatment, average reporting on the outcome variables in the treatment and control group were statistically indistinguishable. While this balance on the outcome variables is indeed reassuring, we also encountered one imbalance (age of the firms). Firms in our treatment group were on average 1.5 years younger than firms in our control group. In Appendix A.2 we show that controlling for age does not affect our estimates.

After treatment, firms in the treatment group reported substantially more wage and more employees. Using a first-difference approach, the estimated ITT on firms' wage reporting is NOK 100,000, which amounts to a 18 percent increase. This provides compelling evidence that wages have previously been underreported and that audits had a deterrence effect increasing compliance. Furthermore, the effect on the reported number of employees is even stronger, with 1.1 employees or a 22 percent increase, which may be taken to suggest that underreporting has taken the form of completely unreported workers, rather than only unreported cash paid to declared workers in addition to their reported wage.

The audits do not seem to have had any effect on the probability of a firm reporting wages on the extensive margin, suggesting that the evading firms have not been completely informal when it comes to reporting workers (or at least that the deterred evasion does not comprise this form of evasion).

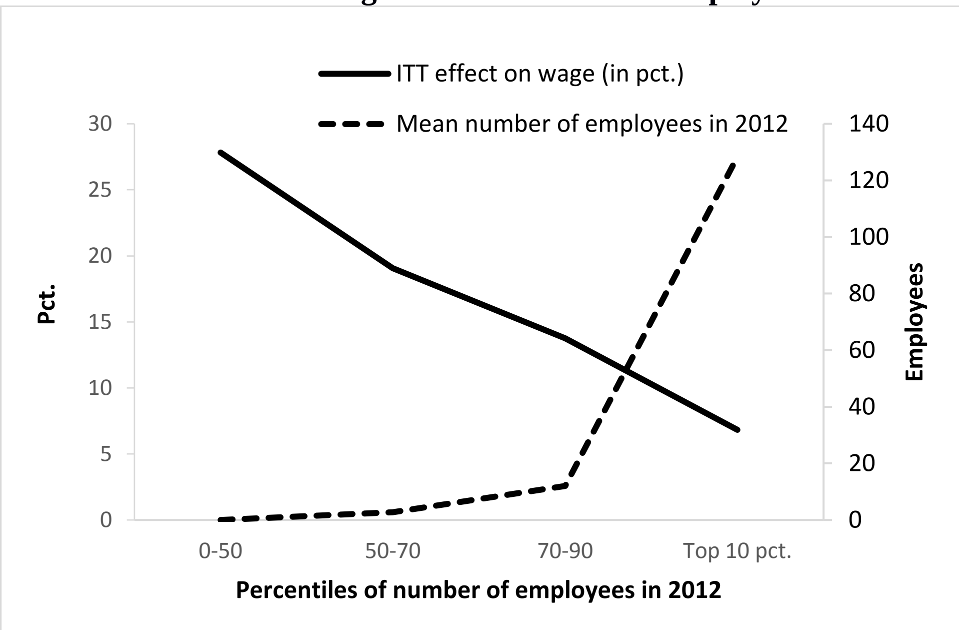
Table 3 Estimated effects of being assigned an audit (ITT)

	Treatment group	Control group	Difference	Robust s.e.
Wage (mill. NOK)				
2012	0.40	0.41	-0.01	(0.04)
2014	0.67	0.58	0.09	(0.06)
Change 2012-2014	0.27	0.17	0.10*	(0.05)
2014 with control for lagged value			0.10*	(0.05)
Number of employees				
2012	3.37	3.38	-0.02	(0.26)
2014	6.20	5.10	1.10*	(0.56)
Change 2012-2014	2.83	1.72	1.12*	(0.52)
2014 with control for lagged value			1.11*	(0.52)
Fraction of firms reporting positive wages				
2012	0.43	0.43	-0.00	(0.02)
2014	0.51	0.50	-0.01	(0.03)
Change 2012-2014	0.08	0.07	-0.01	(0.03)
2014 with control for lagged value			-0.01	(0.03)
N observations	**763**	**1,354**	**2,117**	

Note: The columns denoted Treatment group and Control group include reported wage, reported number of employees and fraction of firms reporting positive wages by treatment and control group in 2012 and 2014. The next column (Difference) provides effect estimates from OLS cf. Eq. (1). Each figure from a separate regression on the given dependent variable and weighted with the number of firms in each stratum as defined in Section 5/Appendix A.0. Main analytic sample cf. Section 4.3. Robust standard errors account for heteroscedasticity. * indicates significance at the 5 percent level.

As noted above, a number of authors have underlined that collusive tax evasion is very hard in large firms, but feasible and economically attractive in small firms (Kleven et al. 2016; Barth and Ognedal 2017). To explore this, we have estimated ITT across sub-groups of firms by their number of employees in 2012. Results are provided in Appendix Table A.1, and we see that the relative effect estimate is largest for the firms with very few employees in 2012, and gradually decreasing as we add firms with more and more employees in 2012 to the sample. While none of the estimates are statistically significantly different, the expected pattern is even more pronounced in Figure 2 where we have split the sample in mutually exclusive groups. Along with the effect estimate (ITT) for each group, we have also plotted the mean number of employees in each group (in 2012). These empirical patterns align surprisingly well with the theoretical arguments of Kleven et al. (2016) that, while collusive tax evasion can be hard to sustain in firms with many employees and accurate business records, evasion can be more easily coordinated in small firms.

Figure 2 The effect estimate is larger for firms with few employees



6.3 Effects of actually being audited (LATE)

As estimates of the effect of being audited, the ITT estimates presented above are too low since half of the firms randomized to be audited did in fact not receive one, and, moreover, since some firms were audited even though they were assigned to the control group.

While the ITT estimates measure effects of being assigned an audit, the LATE estimates measure effects of actually being audited. As we would expect, the LATE estimates, shown in Table 4, are larger in magnitude than our ITT estimates.

The IV estimates show that being audited on average increased the audited firms' wage reporting by NOK 420,000, and the reported number of employees is estimated to increase by 3.6. Using the estimated counterfactual for compliers, assuming a common trend for audited and non-audited firms in the absence of audit, this amounts to increases of 77 percent and 84 percent, respectively. However, these effect estimates are barely marginally statistically significant (p-values at about 10 percent). The fraction of firms reporting wage is estimated to increase by 3 percentage points, from 79 percent to 82 percent, but this estimate is not close to statistically significant at any conventional levels.

The mostly statistically insignificant estimates may relate to the fact that being assigned an audit is a weak instrument of actually being audited. While the correlation between being assigned an audit and actually being audited is positive in all strata, it differs from 0.05 to 1 across strata. Moreover, in almost half of the strata it is not statistically significant at the 5 percent level (t-test), and in only 5 of the 12 strata the F-test statistic exceeds the rule-of-thumb of 10 for adequately strong instruments (Stock et al. 2002). Since we have exactly the same number of instruments as we have endogenous variables (one for each stratum), this implies that we should expect imprecise LATE estimates (Angrist and Pischke 2009, p. 209).[17]

Ignoring the low statistical power, and elaborating on the interpretation of the estimated coefficients, it is important to note that the estimates are only capturing the effect of being audited for the firms that were audited as a result of being assigned an audit ("compliers"), estimated to constitute approximately 45 percent of our sample.[18] These firms are likely to differ from firms that would, or would not, be audited regardless of assignment. This should be taken into consideration when evaluating their external validity. Our LATE estimates can be interpreted as the average effect of audits on the firms that are audited if and only if randomly assigned an audit. This is a relevant measure for the Tax Administration, as these firms are likely to be on the margin of being audited.

For firms that were not audited even though they were assigned to the treatment group ("never-takers"), being audited would not have any effect on their reporting given that the reason they were not audited was that they did not exist or for other reasons were not required to keep a staff register. On the other hand, we would expect audited firms from the control group ("always-takers") to be the most suspicious firms with potentially larger effect of audits than our LATE estimates for compliers. While we are not able to consistently estimate the effect of audit for this group, we see from Table 5 that the increase in wage reporting from 2012 to 2014 is in fact *smaller* (statistically insignificantly so, though) for audited firms in our control group (always-takers) than for audited compliers in the treatment group. This may be

[17] The weak instrument bias (toward OLS), along with too small estimated standard errors, is less of a concern for just-identified estimators (Angrist and Pischke 2009). The limited information maximum likelihood (LIML) estimator can be less biased (and without similarly too small standard errors) when instruments are weak (op. cit.). Our results are virtually unchanged, including standard errors, when we apply LIML instead of 2SLS.

[18] Given monotonicity, the audited firms in our control group are always-takers and the non-audited firms in our treatment group are never-takers. Because the assignment was random, we assume that the share of always-takers (3 percent) and never-takers (52 percent) is equal in the treatment and control group.

taken to suggests that our LATE estimates are not seriously downward biased by some firms in the control group being audited.

Table 4 Estimated effects of actually being audited (LATE)

	Compliers in treatment group (audited)	Compliers in control group (not audited)	IV	Robust s.e.
Wage (mill. NOK)				
2012	0.46	0.43	0.03	(0.36)
2014	0.97	0.52	0.45	(0.51)
Change 2012-2014	0.51	0.09	0.42	(0.26)
Diff 2014 with control for lagged value			0.42	(0.26)
Number of employees				
2012	3.55	3.23	0.32	(1.96)
2014	7.78	3.90	3.88	(2.59)
Change 2012-2014	4.23	0.67	3.56	(2.23)
Diff 2014 with control for lagged value			3.67	(2.16)
Fraction of firms reporting positive wages				
2012	0.49	0.49	0.00	(0.18)
2014	0.82	0.79	0.03	(0.26)
Change 2012-2014	0.33	0.31	0.03	(0.16)
Diff 2014 with control for lagged value			0.03	(0.22)
N observations			**2,117**	

Note: The columns denoted Compliers in treatment group and Compliers in control group include estimated reported wage, reported number of employees and fraction of firms reporting positive wages for the compliers in the treatment and control group in 2012 and 2014. The next column (IV) provides effect estimates from 2SLS with audited dummy interacted with strata dummies, and audited being instrumented by being assigned an audit for each stratum separately; see Eqs. (2) and (3) in Section 5. Each figure comes from a separate regression on the given dependent variable and weighted with the number of firms in each stratum as defined in Section 5/Appendix A.0. Main analytic sample cf. Section 4.3. Robust standard errors account for heteroscedasticity. * indicates significance at the 5 percent level.

Table 5 Estimated average reported wage in 2012 and 2014 for compliers, never-takers and always-takers

	2012	2014	Change 2012-2014
Wage (mill. NOK)			
Never-takers (not audited)	0.35	0.48	0.14
Compliers in control group (not audited)	0.43	0.52	0.09
Compliers in treatment group (audited)	0.46	0.97	0.51
Always-takers (audited)	0.65	0.92	0.27
Number of employees			
Never-takers (not audited)	3.01	4.36	1.35
Compliers in control group (not audited)	3.23	3.90	0.67
Compliers in treatment group (audited)	3.90	7.78	4.23
Always-takers (audited)	4.84	8.33	3.49
Fraction of firms reporting positive wages			
Never-takers (not audited)	0.43	0.36	-0.08
Compliers in control group (not audited)	0.49	0.79	0.31
Compliers in treatment group (audited)	0.49	0.82	0.33
Always-takers (audited)	0.46	0.69	0.23

Note: Reported wage for the compliers in the treatment and control group, as well as for never-takers (firms not audited in treatment group) and always-takers (firms audited in control group). All figures weighted with the number of firms in each stratum, as defined in Section 5/Appendix A.0. Main analytic sample cf. Section 4.3.

7. Conclusion

In many countries, governments are striving to improve the efficiency and effectiveness of the tax administration. A large share of tax administrations' resources are devoted to tax audits and other verification-related activities. These activities vary a lot in their scope and intensity. Understanding the mechanisms of tax evasion and the impacts of different enforcement measures is important both for improving the monitoring and enforcement strategies by tax administrations and, more generally, for improving the efficiency, legitimacy and distributional effects of the tax system.

A vital part of enforcement in modern tax systems is third-party reporting. When functioning at its best, third-party reporting could prohibit tax evasion with minimal enforcement resources. When incentives by third-parties and the taxpayer coincide, however, the preventive effects of third-party reporting are no longer powerful. We produce evidence of collusive tax evasion by employers and employees using a field experiment with firms randomly assigned to on-site audits. In line with theory, we show that on-site audits increase the firms' reporting of wages and number of employees, especially in firms with few employees where frictionless collusion is more likely.

Our findings suggest that third-party reporting cannot always be trusted. Parties can collude to misreport in return for a portion of the tax savings. This is important to consider for tax administrations when allocating their enforcement resources – they need to identify in which situations the third-party can be trusted and in which situations they are likely to underreport. Also, this could have implications for the design of the tax system, taking into account that evasion can lead to labour market distortions and efficiency losses.

Although our results are specific to our setting and even though our findings cannot disentangle the exact mechanisms at work, they highlight that even non-extensive on-site audits could have important deterrence effects. Such benefits of audits should be taken into account when designing an optimal enforcement strategy.

References

Abraham, M., Lorek, K., Richter, F., & Wrede, M. (2017). Collusive tax evasion and social norms. *International Tax and Public Finance, 24*(2), 179-197.

Advani, A., Elming, W., & Shaw, J. (2017). The dynamic effects of tax audits. *IFS Working Paper W17/24*.

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics, 1*(3), 323-338, doi:10.1016/0047-2727(72)90010-2.

Angrist, J., & Pischke, J.-S. (2009). Mostly harmless econometrics: an empiricists guide. Princeton University Press.

Athey, S., & Imbens, G. W. (2017). Chapter 3 - The Econometrics of Randomized Experimentsa. In A. V. Banerjee, & E. Duflo (Eds.), *Handbook of Economic Field Experiments* (Vol. 1, pp. 73-140): North-Holland.

Barth, E., & Ognedal, T. (2017). Tax Evasion in Firms. *LABOUR*, doi:10.1111/labr.12111.

Besim, M., & Jenkins, G. P. (2005). Tax compliance: when do employees behave like the self-employed? *Applied Economics, 37*(10), 1201-1208, doi:10.1080/00036840500109407.

Boadway, R., Marceau, N., & Mongrain, S. (2002). Joint tax evasion. *Canadian Journal of Economics/Revue canadienne d'économique, 35*(3), 417-435, doi:10.1111/1540-5982.00138.

Boning, W. C., Guyton, J., Hodge, R. H., Slemrod, J., & Troiano, U. (2018). Heard it Through the Grapevine: Direct and Network Effects of a Tax Enforcement Field Experiment. National Bureau of Economic Research.

D'Agosto, E., Manzo, M., Pisani, S., & D'Arcangelo, F. M. (2018). The Effect of Audit Activity on Tax Declaration: Evidence on Small Businesses in Italy. *Public Finance Review, 46*(1), 29-57, doi:10.1177/1091142117698035.

DeBacker, J., Heim, B., Tran, A., & Yuskavage, A. (2015a). Once bitten, twice shy? The lasting impact of IRS audits on individual tax reporting. *Journal of Financial Economics, 117*(1), 122-138.

DeBacker, J., Heim, B. T., Tran, A., & Yuskavage, A. (2015b). Legal Enforcement and Corporate Behavior: An Analysis of Tax Aggressiveness after an Audit. *Journal of Law and Economics, 58*(2), 291-324.

Di Porto, E. (2011). Undeclared Work, Employer Tax Compliance, and Audits. *Public Finance Review, 39*(1), 75-102, doi:doi:10.1177/1091142110381641.

Dörrenberg, P., & Schmitz, J. (2017). Tax compliance and information provision - A field experiment with small firms. *Journal of Behavioral Economics for Policy, Vol. 1, No. 1*.

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

Hashimzade, N., Myles, G. D., Page, F., & Rablen, M. D. (2014). Social networks and occupational choice: The endogenous formation of attitudes and beliefs about tax compliance. *Journal of Economic Psychology, 40*, 134-146, doi:http://dx.doi.org/10.1016/j.joep.2012.09.002.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark.(Report). *Econometrica, 79*(3), 651.

Kleven, H. J., Kreiner, C. T., & Saez, E. (2016). Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries. *Economica, 83*(330), 219-246, doi:10.1111/ecca.12182.

Kolm, A.-S., & Bo Nielsen, S. (2008). Under-reporting of Income and Labor Market Performance. *Journal of Public Economic Theory, 10*(2), 195-217, doi:10.1111/j.1467-9779.2008.00358.x.

Kolm, A.-S., & Larsen, B. (2018). Underground activities and labour market performance. *Copenhagen Business School, Working paper;1-2018*.

Kumler, T., Verhoogen, E., & Frías, J. A. (2013). Enlisting employees in improving payroll-tax compliance: Evidence from Mexico. *National Bureau of Economic Research, No. 19385*.

Madzharova, B. (2013). The Effect of a Low Corporate Tax Rate on Payroll Tax Evasion. In C. Fuest, & G. R. Zodrow (Eds.), *Critical Issues in Taxation and Development* (pp. 109-144). Cambridge MA: MIT Press.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, *99*(2), 210-221, doi:https://doi.org/10.1016/j.jdeveco.2012.01.002.
Mittone, L. (2006). Dynamic behaviour in tax evasion: An experimental approach. *Journal of Socio-Economics, 35*(5), 813-835, doi:10.1016/j.socec.2005.11.065.
Nygård, O. E., Slemrod, J., & Thoresen, T. (2016). Distributional Implications of Joint Tax Evasion. *CESifo Working Paper Series, CESifo Group Munich., 5915*.
OECD (2015). *Tax Administration 2015: Comparative Information on OECD and Other Advanced and Emerging Economies*: OECD Publishing, Paris.
Pissarides, C. A., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics, 39*(1), 17-32, doi:10.1016/0047-2727(89)90052-2.
Pomeranz, D. (2015). No taxation without information: deterrence and self-enforcement in the value added tax. *American Economic Review*, doi:10.5167/uzh-136556.
Rubin, D. B., & Imbens, G. W. (2015). Stratified Randomized Experiments. In *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* (pp. 187-218). Cambridge: Cambridge University Press.
Sandmo, A. (2012). An evasive topic: theorizing about the hidden economy. *International Tax and Public Finance, 19*(1), 5-24, doi:10.1007/s10797-011-9185-9.
Santoro, A. (2017). Do Small Businesses Respond to a Change in Tax Audit Rules? Evidence from Italy. *Public Finance Review, 45*(6), 792-814, doi:10.1177/1091142116676361.
Slemrod, J. (2001). A General Model of the Behavioral Response to Taxation. *International Tax and Public Finance*, *8*(2), 119-128, doi:10.1023/A:1011204301325.
Slemrod, J., Collins, B., Hoopes, J. L., Reck, D., & Sebastiani, M. (2017). Does credit-card information reporting improve small-business tax compliance? *Journal of Public Economics, 149*, 1-19.
Stock, J. H., Wright, J. H., & Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics, 20*(4), 518-529.
Telle, K. (2013). Monitoring and enforcement of environmental regulations. *The Journal of Public Economics, 99*, 24.
Thorsager, M., & Melsom, A. M. (2017). Personallisteordningen - motvirker den svart arbeid? *Skatteetatens analysenytt, 02/2017*.
Tonin, M. (2011). Minimum wage and tax evasion: Theory and evidence. *Journal of Public Economics, 95*(11), 1635-1651, doi:10.1016/j.jpubeco.2011.04.005.
Yaniv, G. (1992). Collaborated Employee-Employer Tax Evasion. *Public Finance-Finances Publiques, 47*(2), 312-321.

Appendix A.0 Weighting across strata.

To get a consistent estimator for the population average ITT, we weight by the share of firms in each stratum. This can be done by running OLS on Eq. (1) and weight the estimates and calculate the standard errors (Athey and Imbens 2017) or by transforming the parameter vector as follows (cf. Rubin and Imbens 2015):

$$y_i = \tau \times Treat_i \times \frac{C_{iJ}}{q_J} + \sum_{j=1}^{J} \beta_j \times C_{ij} + \sum_{j=1}^{J-1} \tau_j \times Treat_i \times \left(C_{ij} - C_{iJ} \times \frac{q_j}{q_J} \right) + u_i$$

which gives us a consistent estimator for the population average treatment effect $\tau = \sum_{j=1}^{J} q_j \times \tau_j$, where $q_j = N_j/N$ equals the proportion of firms in each stratum. If treatment effects were constant across strata, one could run the traditional, more convenient and more efficient model controlling for stratum additively:

$$y_i = \tau \times Treat_i + \sum_{j=1}^{J} \beta_j \times C_{ij} + u_i$$

This implicitly weights each stratum with weights proportional to the product of the stratum size (fraction of observations in the stratum) *and* the stratum proportions of treated and control units (probabilities of receiving and not receiving treatment). The motivation is that *ceteris paribus*, a stratum with more variation in treatment (e.g. where there is an equal proportion of treated and control units) provides more information about the treatment effect than a stratum with less variation in treatment (e.g. where only one, or all but one, firm is treated). However, if, as in our case, the fraction of treated units varies a lot across strata and treatment effects are heterogeneous, this will not be a consistent estimator for the population average ITT (cf. Rubin and Imbens 2015).

Appendix A.1 Regression results for different groups based on firms size

In our main analysis, we excluded the 10 percent largest firms in terms of number of employees, since they are presumed less able to engage in collusive tax evasion (Kleven et al. 2016). Here, we show estimated intention to treat effects (ITT) for all firms, as well as when we exclude the 50, 40, 30, 20, 15, 10 and 5 percent largest firms measured by number of employees in 2012 (pre-treatment).

In general, point estimates for reported wages and number of employees are increasing in absolute value with firm size. But as expected from the arguments of e.g. Kleven et al. (2016), measured in relative values (percent of the estimated counterfactual), the effects are larger for the smaller firms. Especially, firms in the 0-50 percentile, e.g. firms that did not report any employees in 2012, seems to have had the largest relative increase. When we zoom in on the smallest firms, sample sizes drop and the estimates are not statistically significant. Furthermore, the effect estimates are not statistically significant when we include the whole population of firms. The lack of precision for the whole population is not surprising, considering the large variation in firm size, as shown in the lower right part of the table.

Table A.1 Estimated intention to treat effects (ITT) of being assigned an audit by firm size

	Percentile (by number of employees in 2012)							
	0-50	0-60	0-70	0-80	0-85	0-90	0-95	0-100
Wage								
Effect estimate	0.064	0.047	0.057	0.083*	0.074	0.103*	0.117*	0.189
Robust s.e.	(0.058)	(0.050)	(0.048)	(0.046)	(0.045)	(0.052)	(0.056)	(0.226)
Implied relative effect	28 %	20 %	20 %	22 %	16 %	18 %	17 %	9 %
Number of employees								
Effect estimate	0.871	0.609	0.684	0.985	0.967	1.117*	1.395*	1.650
Robust s.e.	(0.683)	(0.592)	(0.559)	(0.523)	(0.503)	(0.522)	(0.567)	(0.993)
Implied relative effect	32 %	23 %	24 %	28 %	24 %	22 %	21 %	12 %
Fraction of firms reporting wages								
Effect estimate	-0.005	-0.043	-0.028	-0.015	-0.011	-0.007	0.001	-0.001
Robust s.e.	(0.033)	(0.037)	(0.034)	(0.032)	(0.030)	(0.029)	(0.028)	(0.027)
Implied relative effect	-2 %	-12 %	-7 %	-3 %	-2 %	-1 %	0 %	0 %
Max. number of employees in 2012	0	2	5	10	15	26	46	>1000
Median number of employees in 2012	0	0	0	0	0	0	0	1
Mean number of employees in 2012	0	0.24	0.76	1.62	2.36	3.45	5.12	15.69
St.dev. number of employees in 2012	0	0.58	1.44	2.69	3.81	5.77	9.09	113.38
Mean number of employees with wage>NOK100,000	0	0.11	0.31	0.65	0.91	1.23	1.70	6.97
Observations	1,171	1,406	1,625	1,861	1,992	2,117	2,237	2,355

Note: First difference estimates of the effect of being assigned audit on reported wage, reported number of employees and fraction of firms reporting positive wages. Estimates from OLS cf. Eq. (1). Weighted with the numbers of firms in each stratum as defined in Section 5/Appendix A.0.Estimates from OLS with treatment dummy interacted with strata dummies. Robust standard errors account for heteroscedasticity. * indicates significance at the 5 percent level.

Appendix A.2 Regression results when controlling for age and lagged dependent variable

We recall from Table 3 that the outcome variables are nicely balanced in 2012 (pre-treatment), but as noted, firms in our treatment group were on average younger than firms in our control group. Here we control for firms' age additively to ensure that imbalance in age does not affect our estimates.

Table A.2. shows our main ITT results in column 1 (difference post-treatment) and column 5 (first difference), and compare them to the results when we control for age (column 2 and 6), for lagged dependent variable (column 3) or both (column 4). We see that this does not affect our estimates considerably.

Table A.2 Estimated intention to treat effects (ITT) of being assigned an audit when controlling for firm age and lagged dependent variable

	Difference 2014				First difference 2014-2012	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Main analytic sample (excluding the 10 percent largest firms in 2012)						
Wage	0.088	0.097	0.103*	0.092	0.103*	0.092
	(0.063)	(0.064)	(0.052)	(0.051)	(0.052)	(0.051)
Number of employees	1.102*	1.136*	1.112*	1.016*	1.117*	0.984*
	(0.555)	(0.546)	(0.519)	(0.493)	(0.522)	(0.494)
Fraction of firms reporting any wage	-0.008	-0.011	-0.008	-0.012	-0.007	-0.013
	(0.025)	(0.025)	(0.025)	(0.023)	(0.029)	(0.027)
Observations	2,117	2,117	2,117	2,117	2,117	2,117
Panel B: All firms						
Wage	-0.575	0.177	0.215	0.165	0.189	0.165
	(1.151)	(1.031)	(0.209)	(0.152)	(0.226)	(0.162)
Number of employees	-2.497	0.989	1.676	1.542	1.650	1.536
	(4.709)	(3.907)	(1.101)	(0.929)	(0.993)	(0.915)
Fraction of firms reporting any wage	-0.001	-0.001	-0.001	-0.002	-0.001	-0.004
	(0.023)	(0.023)	(0.023)	(0.021)	(0.027)	(0.025)
Observations	2,355	2,355	2,355	2,355	2,355	2,355
Controlling for:						
Age of firm		X		X		X
Lagged dependent variable			X	X		

Note: Difference and first difference estimates of being assigned to treatment (ITT) on reported wage, reported number of employees and fraction of firms reporting positive wages. Estimates from OLS cf. Eq. (1). Each figure from a separate regression on the given dependent variable and weighted with the number of firms in each stratum as defined in Section 5/Appendix A.0. Main analytic sample in Panel A and all firms in Panel B; cf. Section 4.3. Robust standard errors account for heteroscedasticity. * indicates significance at the 5 percent level.

Appendix A.3 Regression results when controlling for strata additively

The standard way to estimate intention to treat effects (ITT) in experiments with random assignment, is simply to compare differences in overall means across the treated and control groups post-intervention. Unfortunately, and as discussed in Section 5 and Appendix A.0, this approach is not likely to yield consistent estimates in our setting, mainly because effect estimates are heterogeneous across strata (this is confirmed by formal F-tests). The risk-based stratification causes outcome variables to be imbalanced in the overall sample in 2012, which both we and Thorsager and Melsom (2017) find, unless we weight as outlined in Appendix A.0. In our main analysis we have therefore applied the more general model with strata interacted with the treatment dummy, cf. Sections 5 and 6 (and Appendix A.0). Still, here, we show estimated intention to treat effects (ITT) and Local average treatment effects (LATE) in the more restrictive model where we control for strata additively. In general, we see that our estimates do not change much (given the standard errors), though they are generally somewhat lower and with less precision.

Though the imbalance is clearly visible in their plots, Thorsager and Melsom (2017) do not find that the imbalances are statistically significant. The method they apply (difference-in-differences) should not be sensitive to such possible imbalance. Their approach differs from ours as they look at all firms regardless of size and do not weight with the number of firms in each stratum as we have outlined in Section 5/Appendix A.0. They find no statistically significant effects (on the outcome variables available to them - they had more short-term data on e.g. reported turn-over), which is broadly also our ITT findings when we look at all firms regardless of size and do not weight with the number of firms in each stratum; see lower panel (second column) of Table A.3.

Table A.3 Estimated intention to treat effects (ITT) of being assigned an audit and local average treatment effects (LATE) of being audited when controlling for strata additively

	ITT		LATE	
	Control for strata interacted with treatment	Control for strata additively	Control for strata interacted with treatment	Control for strata additively
Panel A: Main sample (excluding 10 percent largest firms in 2012)				
Wage	0.103*	0.098	0.420	0.243
	(0.052)	(0.065)	(0.255)	(0.160)
Number of employees	1.117*	1.052	3.559	2.604
	(0.522)	(0.638)	(2.227)	(1.565)
Fraction of firms reporting any wage	-0.007	-0.041	0.028	-0.102
	(0.029)	(0.026)	(0.155)	(0.065)
Observations	2,117	2,117	2,117	2,117
Panel B: All firms				
Wage	0.189	0.208	0.947	0.505
	(0.226)	(0.333)	(1.177)	(0.804)
Number of employees	1.650	1.257	9.588	3.048
	(0.993)	(1.344)	(8.716)	(3.249)
Fraction of firms reporting any wage	-0.001	-0.033	0.023	-0.080
	(0.027)	(0.023)	(0.123)	(0.057)
Observations	2,355	2,355	2,355	2,355

Note: First difference estimates of being assigned an audit (ITT) and of being audited (LATE) on reported wage, reported number of employees and fraction of firms reporting positive wages. Estimates from OLS (ITT) and 2SLS (LATE) cf. Eq. (1) and Eqs. (2)/(3), respectively. Each figure from a separate regression on the given dependent variable and weighted with the number of firms in each stratum as defined in Section 5/Appendix A.0. Main analytic sample in Panel A and all firms in Panel B; cf. Section 4.3. Robust standard errors account for heteroscedasticity. * indicates significance at the 5 percent level.