The Effects of Labour Migration and Interventions on Tax Compliance

Thomas Lange

Dissertation submitted to the Norwegian School of Economics, for the degree of Philosophiae Doctor (PhD)

Oslo, December 2023

Contents

Acknowledgments

First of all, I would like to thank my supervisors at the Norwegian School of Economics (NHH), Evelina Gavrilova-Zoutman and Floris Zoutman for excellent guidance throughout my PhD. Following my many years outside academia, prior to my PhD journey, their effort was timely and appreciated. I would also like to thank Jarle Møen for facilitating the admission to the PhD program at the NHH. Thanks also to Lars Jonas Andersson at NHH for providing good guidance on machine learning, and support for my notion on its relevance to this project.

This PhD could not have been realised without the funding and support from the Norwegian Research Council and the Norwegian Tax Administration (NTA). But equally important, the many brilliant colleagues in the latter institution. I would like to thank Marcus Zackrisson who paved the way for this project at management level, and Terje Nordli and Monica Bredesen for anchoring the interest among the colleagues engaged in the important work of preventing tax related labour market crime. Research coordinator at the NTA, Torhild Henriksen, deserves a special thanks for facilitating and maintaining research interaction between the NTA and the NHH, and for giving professional advice during the whole period.

I am very thankful for the great talks and substantial contributions to machine learning provided by Nils Gaute Voll. Øystein Olsen and Tore Sjøstedt helped a lot with specifying the data extraction. I would like to thank Hanne Beate Næringsrud for the many comments on specific issues, including interpretation, message, and narratives. Nina Serdarevic, Julia Tropina Bakke, Knut Løyland, Inge Sandstad Skrondal, and Arnstein Øvrum have provided useful comments on earlier drafts as well. I would also like to thank Kari Djupdal, Anders Berset, Andreas Olden, Joakim Døving Dalen, and Terje Dalen for the many fruitful discussions I have had over the years at the NTA. What a great knowledge pool you all are!

A special thanks goes to Anne May Melsom who was appointed my co-supervisor at the NTA and co-authored two of the papers. Not only do I owe you for improving my Stata knowledge to an adequate level, but also for your impeccable understanding of the data, your substantial inquiries, and finally for the considerable effort to the very end. Your help has been invaluable.

I would also like to thank academic staff at other institutions, for their valuable insights and comments. Those are Joel Slemrod at the University of Michigan, Ann Arbor, Steinar Strøm, Andreas Kotsadam and Thor Olav Thoresen at the University of Oslo, and Hamed Saiedi at the Norwegian Business School.

I am grateful for the many conversations and colloquial preparations with peer students at the Stockholm Doctoral Course Program in Economics, Econometrics and Finance (SDPE) jointly by Stockholm University and Stockholm School of Economics, where I undertook most of my

course components. I am equally thankful for the talks and encounters at various conferences with my peer PhD students at the NHH.

A particular gratitude goes to my dear friends at the NTA, namely Hanne Beate Næringsrud, Julia Tropina Bakke, Øystein Olsen, Anders Berset, and Ivana Haakens for being there in challenging times. Friends like You last a lifetime.

I am forever thankful to Rebecka Maria Norman for the continuous, but nevertheless (at least for me) useful discussions on so many topics on statistical inference over the years. I hope our kids, Felicia, Gabriel, and August, were not permanently damaged by nitty gritty talks on standard error clustering or heteroscedasticity. Their patience has been remarkable.

Finally, my heartfelt gratitude to Ann-Kristin Midtskog for new perspectives on compliance, text clarifications, strategic choices, and for your unconditional love and support ("…Og jeg kan ikke miste det uansett hva som skjer").

Oslo, December 2023

Thomas Lange

Introduction

The drivers behind tax non-compliance are diverse. They range from conscious withholding of taxable income and assets to unconscious errors made in the effort of trying to comply with complex declaration standards.

This thesis aims to understand tax compliance and non-compliance. To bring new knowledge to the research literature, I study tax compliance from three different perspectives: The employees, the managers and, finally, the firm.

Since most people generally prefer to have more money rather than less, there will always be a fraction of the taxpayer population trying to evade taxes. A tax system with exemptions, loopholes and complex compliance rules provides ample opportunities for the creative evader. However, these same 'imperfections' also make compliance more difficult for the compliant. Thus, the taxpayer's sense of duty may not be sufficient for compliance in either case.

The Norwegian tax system has since its early origins been based on the idea that trust and obligation, combined with the permanent provision of public goods, will maintain compliance (Brautigam et al., 2008; Holte, 2020). A stable and homogeneous workforce up until the EU expansion in 2004 has most likely made this notion maintainable. However, in 1990, approximately 3.5 percent of the Norwegian population was foreign-born, and this fraction has increased to 14.8 in 2021 (Statistics Norway, 2022). With the EU expansion in 2004, Norway has seen a significant increase in the mobility and heterogeneity of the workforce, with an increasing fraction of labour market participants coming from countries with different taxpayer cultures. Lassen (2007) find that increased ethnic fractionalisation may reduce voluntary compliance. Bastani et al. (2020) study the use of commuter deductions among different immigrant groups in Sweden and find that differences between groups' tax behaviour diminish over time.

Based on a unique dataset from 1,974 randomly audited firms, we have been able to study these and other effects on compliance behaviour of both employees and managers, and, finally how audits and information letters as enforcement strategies affect the firm's compliance obligations.

In the first essay, we analyse the effects of two relevant proxies for cultural background which may affect tax compliance, namely corruption level and conflict exposure in the country of origin (Lange & Melsom, 2022).

Corruption level is measured by the Corruption Perception Index (CPI) (Transparency International, 2019a). CPI is a widely used indicator of corruption in academic literature (Lambsdorff, 1999), and our hypothesis is that the index displays an array of various individual

trust elements towards government, including trust in that the tax collection will yield returns of public goods and not end up in civil servants' pockets. We argue that the CPI Score reflects a person's perception of corruption environment and is relatively stable. We expect the effect of corruption to be negative on compliance.

Conflict, on the other hand, represents more abrupt societal changes and is synonymous with the destruction of trust between ethnic, religious, or other groups, and institutions typically do not recover immediately after a cycle of violence (Collier and Sambanis (2002); Miguel et al. (2011); Bellows and Miguel (2006)). Understood as a trauma, we expect the effect of conflict to also have a negative impact on compliance.

We find associations between conflict exposure prior to migration, and a reduction in probability of compliance, measured by holding a written employment contract, up to 9.9 percent. This result is statistically significant at .01-level. We also find a compliance pattern across residence time groups similar to that of Bastani et al. (2020), as compliance is significantly higher for those with residence time between 5 and 10 years than for those with residence time of more than 10 years. We find no clear effect of corruption level in the home country, but managers are slightly less likely to hold a contract than other occupational groups when we controlled for business-specific characteristics.

As managers have a bigger influence over firm compliance than their employees, particularly under a third-party reporting regime, this negative compliance effect calls for a deeper assessment of determinants of management compliance.

The second paper identifies drivers of tax compliance among firm managers. Managers are exposed to a tax administration through more frequent reporting liabilities than employees and are also responsible for third-party reporting on behalf of their employee workforce. Thus, while cultural factors such as corruption and armed conflict may provide some explanatory power of manager compliance, we explore whether other mechanisms not previously studied, are at play. Variable selection is a challenge when studying managers' tax compliance, because they may be affected by several reporting liabilities not previously studied. We overcome this challenge by using two different variable selection strategies; one based on previous compliance literature, and one based on two different machine learning models, namely a Genetic Algorithm (GA) and a Least Absolute Shrinkage and Selection Operator (LASSO), to select among variables suggested by tax auditors in the Norwegian Tax Administration (NTA). This twofold strategy allows us both to test 'common' independent variables from other compliance areas and to explore a magnitude of reporting variables not previously tested scientifically. The use of machine learning in other economic areas such as stock market price predictions (Sable et al., 2017), financial asset portfolio selection (Sefiane & Benbouziane, 2012), or determination of real estate prices (Del Giudice et al., 2017) has given robust findings. Such

models have various applications in variable selection problems (Broadhurst et al. (1997); Tolvi (2004); Cateni et al. (2010); Liu and Ong (2008)) but have to our knowledge not previously been used in the field of taxation.

Tax auditors are a source of knowledge, and when variables in an audit program are selected based on this knowledge, and firm data is collected on a random basis, there is ample room for supervised machine learning methods. Testing such models can, if properly specified, provide additional robustness checks, but the models have limitations concerning causal inference (Pearl, 2018). However, when the variable selection they perform can be supported by economic intuition, there is merit to their application in this context as well. This is also what we demonstrate in the second paper.

We find a negative association between manager compliance and armed conflict exposure of their workforce. Older managers are less inclined to be compliant, likely due to a learning effect of both tax system loopholes and audit frequency. Moreover, the use of an external accountant and a salary system influences manager compliance positively. We find no effect of home country corruption level. These findings call for a better understanding of the drivers behind manager tax compliance, and an investigation into how the firms they manage respond to tax enforcement strategies implemented by the tax authorities. This is the topic of the third, and final paper.

We study and compare the effect of two different enforcement strategies on the firms' payroll tax remittance. An assessment of the effects of different strategies may help inform resource allocation decisions on which strategy to use. The payroll tax is levied on employers on behalf of their employees as part of the financing of the National Insurance Scheme and contributes largely to the Norwegian tax revenue. In 2021, the payroll tax equalled 14.2 percent of total tax revenue, or 22.9 percent of the taxes levied on firms only (Norwegian Ministry of Finance, 2022). In an experimental setting involving two treatments, audits and information letters respectively, we find that treated firms increase their payroll tax remittance following the treatment.

From a total population of 30,961 firms in different labour-intensive businesses in the Norwegian economy, we randomly drew a sample of 1,974 firms receiving an on-site audit, and another sample of 8,000 firms receiving a standardized information letter. Thus, unlike most other contributions, we study the effects of two different enforcement strategies on the same population and in the same institutional setting. Firms subject to audit, letter or no treatment operate in the same tax environment and are liable to the same set of reporting obligations. The setup is similar to that of Boning et al. (2020) and Bjørneby et al. (2021), but

differs in that firms in our sample were not selected based on risk scores.

Audits and letters are two very different compliance instruments, with different cost functions for the Tax Administration, and different response mechanisms on the receiving end. The average cost of audits in this experiment was NOK 12,740 per audit, whereas the average cost of the letters was negligible. On-site audits are more intrusive than impersonalised information letters. While a scheduled visit from a tax auditor typically evokes a feeling of being subject to disclosure of prospective errors, a letter stating the firm's reporting liabilities seems more like a reminder of general rules. One may therefore expect that a change in behaviour succeeding an audit originates in a sense of enforcement, whereas a prospective change following an information letter may be conceived as more 'voluntary.'

In the treatment year of 2018, an audit would increase the average firm's payroll tax remittance by 48.14 percent more than the non-treated firm, whereas the corresponding figure for the letter treatment is 13.54 percent. On average, we find that the audit effect on the payroll tax increment from 2017 to 2018 is NOK 2,265 per audit, and the letter effect is NOK 289 per letter. Hence, the administrative cost of the audit exceeds the revenue it generates, whereas the opposite is true for the letter treatment. However, there may also be other effects from both treatments, like general deterrence effects and network effects (Boning et al., 2020), that may increase the benefits. Finally, an electronic letter is easily scalable to a larger population, which, all else equal, will increase the revenue further.

The results are representative of a population of labour-intensive firms in an advanced tax reporting environment, by international standards. Thus, the external validity of the results is not restricted to the Norwegian context as such, as one can assume that these effects will be reproduced in several advanced, Western tax jurisdictions with gold standard system of information reporting. The results also suggest that tax authorities may test and compare enforcement strategies in other sectors as well.

References

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A Theoretical Analysis. *Journal of Public Economics*, *1*(3-4), 323-338.

Bastani, S., Giebe, T., & Miao, C. (2020). Ethnicity and tax filing behaviour. *Journal of Urban Economics*, *116*, 103215.

Bellows, J., & Miguel, E. (2006). War and institutions: New evidence from Sierra Leone. *American Economic Review*, *96*(2), 394-399.

Bjørneby, M., Alstadsæter, A., & Telle, K. (2021). Limits to third-party reporting: Evidence from a randomized field experiment in Norway. *Journal of Public Economics*, *203*, 104512.

Boning, W. C., Guyton, J., Hodge, R., & Slemrod, J. (2020). Heard it through the grapevine: The direct and network effects of a tax enforcement field experiment on firms. *Journal of Public Economics*, *190*, 104261.

Brautigam, D., Fjeldstad, O.-H., & Moore, M. (2008). *Taxation and state-building in developing countries: Capacity and consent*. Cambridge University Press.

Broadhurst, D., Goodacre, R., Jones, A., Rowland, J. J., & Kell, D. B. (1997). Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry. *Analytica Chimica Acta*, *348*(1-3), 71-86.

Cateni, S., Colla, V., & Vannucci, M. (2010). Variable selection through genetic algorithms for classification purposes. Proceedings of the 10th IASTED International Conference on Artificial Intelligence and Applications, AIA,

Collier, P., & Sambanis, N. (2002). Understanding civil war: A new agenda. *Journal of Conflict Resolution*, *46*(1), 3-12.

Del Giudice, V., De Paola, P., & Forte, F. (2017). Using genetic algorithms for real estate appraisals. *Buildings*, *7*(2), 31.

Engström, P., & Hagen, J. J. E. E. R. (2017). Income underreporting among the self-employed: a permanent income approach. *92*, 92-109.

Holte, H. C. (2020). Å bygge tillit er viktig, men kontroll trengs. *Dagens Perspektiv*. https://www.dagensperspektiv.no/2020/a-bygge-tillit-er-viktig-men-kontroll-trengs

Kolm, A.-S., & Larsen, B. (2019). Underground activities and labour market performance. *International Tax and Public Finance*, *26*(1), 41-70.

Lambsdorff, J. G. (1999). Corruption in empirical research: A review. *Transparency International, processed*, *6*.

Lange, T., & MELSOM, A. M. (2022). Påvirker oppholdstid i Norge etterlevelse hos utenlandske arbeidstakere? *Samfunnsøkonomen*, 28.

Lassen, D. D. (2007). Ethnic divisions, trust, and the size of the informal sector. *Journal of Economic Behaviour & Organization*, *63*(3), 423-438.

Liu, H.-H., & Ong, C.-S. (2008). Variable selection in clustering for marketing segmentation using genetic algorithms. *Expert Systems with Applications*, *34*(1), 502-510.

Lyssiotou, P., Pashardes, P., & Stengos, T. (2004). Estimates of the black economy based on consumer demand approaches. *The Economic Journal*, *114*(497), 622-640.

Medina, L., & Schneider, F. (2017). Shadow economies around the world: New results for 158 countries over 1991-2015.

Miguel, E., Saiegh, S. M., & Satyanath, S. (2011). Civil war exposure and violence. *Economics & Politics*, *23*(1), 59-73.

Norwegian Ministry of Finance, T. (2022). *Prop. 1 LS (2021–2022) For budsjettåret 2022 — Skatter, avgifter og toll 2022*.

Nygard, O. E., Slemrod, J. B., & Thoresen, T. O. (2016). Distributional Implications of Joint Tax Evasion.

Pearl, J. (2018). Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv preprint arXiv:1801.04016*.

Pissarides, C. A., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, *39*(1), 17-32.

Sable, S., Porwal, A., & Singh, U. (2017). Stock price prediction using genetic algorithms and evolution strategies. 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA),

Schneider, F., & Buehn, A. J. O. E. (2018). Shadow Economy: Estimation Methods, Problems, Results and Open questions. *1*(1), 1-29. https://doi.org/10.1515/openec-2017-0001

Sefiane, S., & Benbouziane, M. (2012). Portfolio selection using genetic algorithm.

Slemrod, J. (2019). Tax compliance and enforcement. *Journal of economic literature*, *57*(4), 904-954.

Statistics Norway, T. (2022). www.ssb.no

Tolvi, J. (2004). Genetic algorithms for outlier detection and variable selection in linear regression models. *soft computing*, *8*(8), 527-533.

Transparency International. (2019). *Corruption Perceptions Index*

Does Residence Time in Norway Affect Tax Compliance?[1]

Thomas Lange[2]

Anne May Melsom[3]

Abstract

Increased labour immigration over the past 15 years from countries with various degrees of trust in tax administrations makes it relevant to look at whether some cultural factors can affect compliance with tax regulations, and whether residence time in Norway affects this in any way. We use data from randomised audits carried out by the Tax Administration and find a non-linear effect of residence time on tax compliance. Furthermore, we find a negative association between armed conflict exposure in the home country, prior to migration to Norway and tax compliance, but no significant effects from corruption level of the home country.

JEL Classification: H2; H26; J6; J61

[1] Thanks to Evelina Gavrilova-Zoutman, Knut Løyland, Steinar Strøm and Floris Zoutman for useful comments on earlier drafts of this paper. A Norwegian version of the paper was peer-reviewed and published in Samfunnsøkonomen 2022, vol.1 pp. 28-39.

[2] Corresponding author, Norwegian School of Economics.

[3] Norwegian Tax Administration

1. Introduction

In a time of globalisation and the international exchange of labour supply, the question of how individual characteristics or different moral standards in the labour force, may influence tax compliance has gained renewed relevance. Studies on corruption norms, tax morale and compliance behaviour help to inform modern tax administrations in their understanding of non-compliance and appropriate mitigation.

Since the expansion of the EU, Norway has seen a significant increase in the number and mobility of foreign workers. More than 300,000 labour immigrants have come to Norway since 1990 (Statistics Norway, 2022). Due to the strong increase in labour immigration to Norway, compliance of the 'new' workforce also has relevance for the government's finances: To secure tax revenue over time, compliance with the tax system must also be sustained.

In this paper, we use data from The Norwegian Tax Administration's (NTA) randomised audit program and ask whether home-country institutional quality, as measured by corruption level and armed conflict exposure, affects tax compliance in Norway, and if so, whether the effects diminish with residence time.

Some previous studies have addressed the question of the relationship between migration, residence time, and compliance, Bastani et al. (2020) being a recent contribution. They study the use of commuter deductions among different immigrant groups in Sweden and find that the use of commuter deductions among labour immigrants increases with their residence time. However, the study does not necessarily show the incorrect use of deductions, but rather the lack of claiming legitimate deductions, which is not considered a compliance breach. There is, to our knowledge, not much empirical evidence demonstrating different population groups' tax compliance behaviour in a taxpayer environment with high degrees of third-party reporting.

Our paper seeks to achieve a better understanding of cross-border influence on tax compliance by analysing how factors originating from an individual's home country can transcend international borders and impact their compliance behaviour in the host country. By exploring the impact of the home-country institutional environment on tax compliance in Norway, our study broadens the understanding of relationships between an individual's past socio-political environment and their present tax compliance behaviour in a new country.

The incorporation of the institutional context may extend the existing understanding of how corruption levels and armed conflict exposure, which are inherently rooted in the cultural and institutional fabric of an individual's home country, may shape their compliance tendencies. As such, we seek to address compliance beyond individual characteristics, as we shift the focus

to include broader systemic and environmental factors such as corruption level and armed conflict exposure.

In 2018, the NTA undertook randomised audits focused on salary reporting compliance. The data used in this paper originates from these audits, which form part of a broader reorganization aimed at enhancing systematic knowledge about financial crime and overall compliance within society. These audits, which involve the random selection of candidates from a defined target group, aim to generalize the findings across the entire target group, thereby enabling insights into numerous taxpayers, not solely those audited.

The objective of the random audit program is to provide insights into areas and taxpayers that risk-based audits might not systematically cover. The approach of employing randomised audits, involving the random selection of audit subjects, ensures a representative sample and eliminates the need to correct results for selection bias. The data collection covers a one-year period in 2017. To analyse the relationship between residence time and compliance, considering enterprise and NTA auditor characteristics, we use an OLS fixed effects model with various control variables. This approach aims to ensure a robust analysis of the relationship between the cultural background factors, residence time, and compliance, considering a range of contributing factors.

We find a statistically significant negative effect of the individual's exposure to armed conflict, but no significant effect from the level of corruption in the home country. One interpretation of this result is that the traumatic experiences and disruption caused by armed conflict may lead to a disruption of institutional trust followed by a decreased focus or capacity to fulfil tax obligations. The stress and disarray resulting from conflict may negatively impact an individual's ability to adhere to tax regulations, unlike the impact of the corruption environment, which may not be directly translated into individual compliance behaviour, as it may be perceived as an institutional issue distinct from personal obligations. Whereas previous exposure to armed conflict may generate psychological distress and emotional trauma that affects an individual's mental state, potentially leading to reduced attention towards meeting tax obligations, corruption may not evoke the same emotional or psychological response at an individual level, thereby lacking a direct impact on compliance behaviour.

We also find greater compliance among those with residence time in Norway of 5 to 10 years than those exceeding 10 years. Immigrants who have been in Norway for 5 to 10 years may have reached a stage where integration peaks, adhering to local laws and regulations, including tax liabilities. However, those exceeding 10 years might face different circumstances, potentially reaching a point where assimilation might plateau or decline, caused by e.g. learning certain loopholes in the tax system or gaining more knowledge of actual audit

probability. Another explanation of this variance could be that individuals in the 5 to 10-year period might view their stay as more temporary or transitional, possibly prompting stricter adherence to regulations to ensure successful settlement.

The findings offer policy insights for countries receiving labour immigrants, particularly regarding the integration of immigrant populations into the tax system. Understanding the impact of corruption and armed conflict exposure on compliance behaviour informs policymakers about the necessary measures to facilitate immigrant integration and foster tax compliance. These insights guide the formulation of policies that account for the socio-cultural backgrounds of immigrant populations and seek to bridge the gaps arising from differences in institutional quality.

2. Related Literature

One opinion from the research literature is that compliance in society increases when the individual is exposed to the norms and laws that apply when encountering different bodies of public administration (Alm, 2019b; Torgler, 2014). Traditionally, the Norwegian tax system has been based on the idea that trust and obligation, in combination with a stable supply of public welfare goods, will maintain compliance (Brautigam et al., 2008; Holte, 2020). However, increased ethnic fractionalisation may reduce voluntary compliance due to varying degrees of social structures and institutional trust (Lassen, 2007).

Some previous studies have focused on the establishment of institutions like large taxpayer offices and anti-corruption agencies (Baum & Gupta, 2017), and institutional quality and the degree of judicial efficiency (Damania et al., 2004) to explain variations in compliance in other areas of the academic literature.

Studies on corruption norms and tax morale have received increased attention recently (Jahnke & Weisser, 2019). Corruption erodes the perception of fairness in the tax system (Alesina & Angeletos, 2005) and undermines trust in public institutions (Clausen et al., 2011), whereas the breakdown of formal institutions and weakened enforcement during conflict can push individuals towards informal employment or business practices (Ballentine & Sherman, 2003). Armed conflict also tears down the social contract between citizens and the state (Azam & Mesnard, 2003). A recent contribution by Galletta and Giommoni (2023), utilizing historical administrative records from Italy following World War I, finds that exposure to war violence has a negative effect on tax compliance.

Changes in compliance due to temporary or permanent demographic changes have, to our knowledge, not been sufficiently discussed in the research literature. There may be several reasons for this. Rapid demographic changes in a population are often caused by immigration. Some studies look at the effects on tax revenue from migration. However, they focus either on

the macro-income effects of immigration (Harding and Mutascu (2016)) or the effects of migration on tax rates for high-income individuals (Young and Varner (2011); Kleven et al. (2014)). None of these studies discuss the effects on compliance from demographic changes due to migration.

Since Torgler (2007) and Torgler, Demir et al. (2008) studied multicultural differences in compliance with the tax system, research on tax morale is usually concerned with both the underlying culture as a cause of tax evasion (Cummings, Martinez-Vazquez et al. 2009) and, slightly conversely, why tax morale in Western countries is so high, given the low probability of audits and sanctions (Frey and Torgler 2007). DeBacker, Heim et al. (2012) find that American companies with owners from more corrupt countries avoid more tax in the US. Common to these empirical approaches is either the use of questionnaires or the use of risk-based audit data, where the latter always involves some kind of bias. We use randomized audits.

3. Data

3.1 Random Audits

In cooperation with (municipal) tax collectors in 2018, the NTA conducted randomised audits covering the compliance area of salary reporting. The findings in this article are based on the data from these audits. The audits are part of a larger reorganisation, in which the NTA seeks to organise parts of the audit activity for more systematic knowledge building of financial crime and general compliance in society. As a part of this restructuring, the NTA has, to a greater extent than before, introduced randomised audits as an integrated part of its activities. Randomised audits imply that audit candidates are randomly selected from a defined target group so that findings from the audits can be generalised to the entire target group from which the audit candidates are selected. With relatively few audits, the NTA gains knowledge about many taxpayers, not just those audited.

The objective of the random audit program is to learn more about areas and taxpayers where risk-based audits cannot generate systematic knowledge. However, the execution of randomised and risk-based audits is different. Risk-based audits are often more extensive, and the audit actions are largely adapted to the individual enterprise to uncover tax evasion. Thus, this results in a sample selection bias absent in the randomised audit sample. Unlike risk-based audits, randomised audits must follow the same audit actions and be carried out on a large scale to produce results with external validity. Therefore, they are often less extensive and better suited to disclose smaller errors and deviations. The caveats of random audits thus remain, i.e. random audits have the disadvantage that the non-risk-based assessment leads to a more standardized and less efficient approach of the audit as auditors may not have specialized knowledge or expertise in the specific industry or compliance issues they

encounter during the audit. Since the selection of audit targets is done randomly, auditors may have limited understanding of the specific circumstances of the taxpayer and to identify potential areas of non-compliance. Audit selection based on an auditor's risk perception is often perceived as a motivational factor. This factor may be reduced when conducting standardised random audits. All these features of random audits are likely to contribute to an underestimation of compliance in the sample since a smaller number of cases are likely to be disclosed.

While the audit targets in the sample have been randomly selected, the audit targets are not randomly assigned to auditors. There may be systematic allocation of more complex enterprises or enterprises assumed to have a higher risk of non-compliance to more experienced auditors. More experienced auditors may find more mistakes than less experienced ones. This may lead to bias in the data. If companies with foreign managers are considered to be more complex or high-risk enterprises and thus assigned to more experienced authors, an observed difference in compliance may be due to the editor and not company characteristics. We thus account for auditor fixed effects in our model, but we find that this selection bias has only minor effects on some of our explanatory variables.

3.2 Population and sample

The target population for the audits includes enterprises that reported employees and other information in the monthly reporting (a-melding) for 2017. The following delimitation criteria were used to generate the target population.

Table 0: Target population of enterprises.

1. have more than NOK 100,000 in turnover, total operating revenue of NOK 200,000 or more and/or
2. have combined salary expenses of > NOK 500,000
3. have between 5 and 20 employees on average per month
4. have more than NOK 100,000 under item 4500 (subcontractors) and 6700 (Accountancy and consultancy services, etc.) in 2016
5. operate in labour-intensive industries
6. are either a Private limited liability company (AS), Sole proprietorship (ENK) or Norwegian registered foreign company (NUF)
7. have NACE codes where more than half of the enterprises have foreign workers

Note: Conditions 1-7 are conjoint unless otherwise specified.

These delimitation criteria generate a target population totalling 30,961 enterprises. A random and stratified sample from this target population has been selected for audit. The sample is stratified by industry to ensure that all industries in the target population will be represented in the sample, and to make it possible to draw more enterprises from industries with a high risk of non-compliance. A total of 1,974 enterprises were audited. See Appendix 1 for an overview of the sectors included in the target population.

The random audits include a series of topics with several questions, checkpoints and audit actions. The majority of these concern the *enterprise* (kept accounts, kept salary accounts, documented time use, tax deductions, kept staff registers, etc.) In addition, documentation on employment contracts, timesheets, overtime payments, holiday pay, etc. was collected for up to 5 randomly selected *employees* per enterprise. The employees are selected based on registered employment relationships in the monthly reporting for 2017 and based on the month in 2017 where the number of registered employees peaked. For enterprises with more than 5 registered employees, up to 10 are randomly selected and the responsible auditor in the NTA (hereafter auditor) should audit the first five from a list in random order. For enterprises with five or fewer registered employment relationships, all are selected for audit.

The collected information on these employees comprises the sample in this analysis. In addition, we have obtained information from the Tax Administration's data warehouse about the rest of the employees within each enterprise and linked this to the Corruption Perception Index (Transparency International, 2019b) and information about whether they come from countries undergone armed conflict (Gleditsch et al., 2002).

The data is cross-sectional, which has some empirical shortcomings. Since we study a single point in time, we have limited ability to analyse temporal changes, trends, or causal relationships between employee compliance and our independent variables over time. The data alone cannot capture the dynamics and evolution of the variables, and so the ability to draw causal inference is equally limited.

Furthermore, it is likely the results suffer from omitted variable bias, where factors such as legal and regulatory factors, tax system design or education may influence compliance levels. Finally, since we have no information on individual-level time variance, we cannot control for individual heterogeneity or account for time-invariant unobserved factors that could affect compliance. Thus, more research is needed to establish causal relationships.

Because randomised audits involve the random selection of audit subjects, the data collection provides a representative sample. Therefore, it is not necessary to correct the results for selection bias. The data is collected for one year (2017). The choice of method consists of finding and specifying the regression model best suited to study our data and the correlations we are interested in.

4. Empirical Strategy

4.1 How do we measure employee compliance?

We are concerned about the part of an employment relationship between employer and employee where the latter "chooses" to comply or not. It is unlikely that some employees will ask the employer to evade taxable income and demand the surplus since deducting tax in a

third-party reporting scheme is the employer's responsibility. Bjørneby et al. (2018) argue that non-compliance concerning third-party pre-filled tax return items is due to deliberate evasion. The reason for this is that pre-filled information provides the Tax Administration with an effective tool for automatic checks by comparing information from the tax return with the monthly reporting from the respective taxpayer's employer. This reasoning coincides with other contributions such as Abraham et al. (2015), Nygard et al. (2016) and Kolm and Larsen (2019).

When measuring the scope of tax non-compliance, we typically face the challenge of separating the wilful understatement of tax liability from the inadvertent (Gerxhani (2004); Schneider and Enste (2000); Alm, Martinez-Vazquez et al. (2004); Feld and Schneider (2010)). Schneider (2012) seeks to define the shadow economy through different types of informal economic activity that are deliberately concealed from taxation. Our definition is somewhat narrower; we do not want to say anything about the motivation behind non-compliance. Our data does not give a sufficient basis for differentiating between deliberate evasion and unintended errors. The different contributions from Schneider et al. use an adapted macro model for estimating informal economic activity in various countries (e.g., Schneider and Buehn (2018); Medina and Schneider (2017). For critical evaluations of the model, see Breusch (2005) and Feige (2016).

We argue that an employment relationship requires some predetermined attitudes formed by experience from other employment relationships. From the employees' perspective, formalising the salary level, duration of the assignment and content of the work will usually require a written contract. It should be in the employee's interest to have a written employment contract independent of what motivation the employer must have for not offering one, since lack of a written contract will place the employee in a weaker position. Lack of an employment contract can either be due to a lack of knowledge that this is required in Norway or be an indication of unlawful tax evasion in cooperation with the employer (Bjørneby et al. (2018)). It may be that employees who receive employment contracts choose to stay longer in Norway, but we are mainly concerned about whether compliance and residence time correlate, rather than causal inference. We measure compliance with a dichotomous response variable that takes the value 1 if the employee has a contract with the employer, and 0 otherwise.

A caveat remains, however. Our compliance variable covers the cases that are legal and the cases that are illegal and were disclosed during the audit. A group missing is the group of unregistered, non-discovered employees in the audited firms. We find that there is no viable solution to this data measurement issue. Since unregistered employees by default cannot hold a written employment contract, there is a perfect correlation with non-compliance. Thus, depending on the number of omitted observations, which are not measurable, the exclusion of this group in the data set will lead to overestimation of compliance. However, as registered

employees also entail other opportunities for profitable fraud through welfare benefits, the question is how large this group is, and thus how likely their exclusion will drive the results. Nevertheless, our ambition in this paper is to study the compliance of *registered* employees in the target population of 30,961 enterprises in Norway.

4.2 Variables

We are interested in how residence time affects compliance when controlling for institutional variables such as corruption and conflict exposure. The most important independent variable in our analysis is residence time in Norway. We also want to explore correlations between residence time and the probability of holding a written employment contract. We measure net residence time in Norway throughout a lifetime based on migration and registration dates in the National Population Register. For Norwegian citizens who have not lived abroad at any time, the residence time will be equal to the employee's age.

Age itself is known to be associated with tax compliance, and the research results point in the direction of a positive correlation between age and tax compliance (Hofmann et al., 2017). The influence of residence time on tax compliance may thus have an alternative interpretation, which may be unrelated to the concept of cultural adaptation of foreigners to tax compliance over time, namely that older employees may have accumulated more experience and knowledge regarding tax laws and regulations over their working years. This increased familiarity with the tax system may lead to a higher level of tax compliance as they understand their obligations and the potential consequences of tax avoidance (Eriksen & Fallan, 1996). Conversely, younger employees who are less experienced or newly entering the workforce may be less aware of tax requirements, leading to unintentional non-compliance. The elderly are observed to have stronger tax morale than young people, and Nordblom and Žamac (2012) explain why this may be an age rather than a cohort effect through social psychological factors. They find that personal norms may evolve over the life cycle due to past behaviour and the attitudes of peers (normative conformity).

Age can also be associated with different income levels and financial situations. Older employees who have been in the workforce for a longer period may have higher incomes and accumulated wealth compared to younger employees. Due to fewer budget constraints, it is easier for them to afford to be tax compliant. Earning more income than younger generations gives them financial freedom which young citizens may lack (Kirchler, 2007).

Attitudes towards tax compliance can also vary with age, but through this mechanism, the results seem more ambiguous (Riley & Riley Jr, 1989). Older employees who have a longer history of paying taxes and being part of the formal workforce may have a stronger sense of civic duty and responsibility towards tax obligations. They may be more inclined to comply with tax laws due to social norms, peer pressure, or a desire to maintain a positive reputation.

However, in a given cohort, shared life histories can lead to the development of personal, social, and societal norms that either promote or undermine tax compliance. Consequently, the correlation between higher age and tax compliance can be observed in both directions. While some cohorts may exhibit an increase in tax compliance as individuals age and adopt norms that prioritize tax honesty, other cohorts may experience a decline in tax compliance as evolving norms erode the importance of tax honesty. Therefore, higher age can be associated with both higher and lower levels of tax compliance depending on the specific cohort and the norms that shape their behaviour (Hofmann et al., 2017).

These studies utilize different age bins, suggesting no linear relationship between age and tax compliance. Neither do we assume a linear relationship between residence time and compliance, and divide the sample into three different residence groups: 0-5 years, 5-10 years and > 10 years. This is the same intervals used by Bastani et al. (2020). As a robustness test, we have estimated compliance across all residence year dummies, but find no linear relationship, cf. Appendix Figure A3.

We also include two central, institutional variables from the country of origin. The Corruption Perception Index (CPI) scores and ranks countries by their perceived levels of public sector corruption, as viewed by experts and business leaders (Transparency International, 2019a). This is a compiled index of 13 corruption surveys and assessments, collected by a variety of reputable institutions. CPI is an extensively used indicator of corruption in the academic literature (Lambsdorff, 1999).

CPI Scores reflect the corruption environment in the home country and are perceived as a more 'stable' cultural proxy than e.g. armed conflicts (Seleim & Bontis, 2009).

High levels of corruption can erode the perception of fairness in the tax system (Alesina & Angeletos, 2005). When individuals believe that taxes are being misappropriated or used for corrupt purposes, they may feel less inclined to comply with their tax obligations. The perception that their tax payments will not be used for the benefit of society can diminish their motivation to comply with tax obligations. Corruption also undermines trust in public institutions (Clausen et al., 2011). When citizens lack trust in their government's ability to effectively combat corruption, they may become sceptical about the proper use of tax revenues.

In a corrupt environment, where economic activities are conducted outside the reach of formal tax regulations, individuals may opt for informal employment or business practices to avoid corruption-related challenges and perceived unfairness. Employees may feel compelled to engage in bribery or other illegal practices to reduce their tax burden or expedite bureaucratic processes. This non-compliant behaviour can be driven by the perception that compliance with tax obligations does not offer equal benefits or opportunities. When corruption permeates tax

administration and enforcement agencies, compliance efforts may be undermined, leading to lower deterrence and a reduced likelihood of penalties for non-compliance. This weak enforcement environment can further diminish employees' motivation to comply with tax obligations (Leitao, 2016).

The other variable is constructed from the Uppsala Conflict Data Program on armed conflicts in countries and areas since 1946 (Gleditsch et al., 2002)). The threshold for conflict in this dataset is set to 25 armed-conflict fatalities per year. In line with Miguel et al. (2011), we define conflict exposure by whether the employee migrates from a country that has experienced armed conflict in the last 25 years before moving to Norway. A war trauma can have individual, long-term effects throughout a lifetime, and the trust one had in various government functions, such as a tax administration, before the outbreak of war typically takes a long time to rebuild. Thus, we say that conflict exposure earlier than 25 years before migration to Norway does not affect compliance. In those cases, it will only be the corruption level that may affect compliance.

Armed conflict infers a social trauma in society in that it destroys the trust between ethnic, religious or other groups, and institutions typically do not bounce back immediately after a cycle of violence (Collier and Sambanis (2002); Miguel et al. (2011); Bellows and Miguel (2006)). Armed conflict often gives rise to an expanded informal economy, where economic activities occur outside the formal sector and are not regulated by tax laws (Ballentine & Sherman, 2003). The breakdown of formal institutions and weakened enforcement during conflict can push individuals towards informal employment or business practices. As a result, tax compliance among employees may decline as they operate in the informal sector to cope with the circumstances of conflict (Litina & Palivos, 2016).

Armed conflict erodes the social contract between citizens and the state (Azam & Mesnard, 2003). When individuals perceive their government as unable or unwilling to provide security and stability during conflict, their trust in state institutions may diminish. This reduced trust can manifest in a decreased willingness to comply with tax obligations, as individuals may view tax payments as further contributions to a system that fails to protect or support them.

In other words, we assume that armed conflict typically represents a "trauma" that in the medium term destroys institutions that naturally connect the individual to the public administration sector (typically the tax administration).

Gleditsch et al. (2002) measure wars or armed conflicts between well-defined groups. However, in some countries, violence, and hence trauma for citizens, may derive from criminal organizations such as gangs. Ali et al. (2013) utilize data from Kenya, Tanzania, Uganda, and South Africa, and find that payments to non-state actors and criminal gangs in exchange for

protection, tend to reduce individuals' tax-compliant attitude. Unfortunately, we do not have information on criminal gangs or other non-state actors. As such cases are not measured, the potential effects of armed conflict exposure may be underestimated.

4.3 Models

We wish to shed light on the relationship between residence time and compliance, controlled for enterprise characteristics and NTA auditor characteristics. Furthermore, we include several control variables in the regression.

We believe that a linear regression model is best suited even if the dependent variable is binary, and econometric literature suggests using Probit models (Wooldridge, 2010). True, Probit models provide different results than linear models when the mean of the dependent variable is over 0.95 or under 0.05. This is not the case with our data. Furthermore, it is not as easy to include controls for enterprises and auditors (that have many different values) in the Probit model, and our model contains dummy variables making the coefficients of these difficult to estimate and interpret (Caudill, 1988). Therefore, we use linear regressions as the main models. Nevertheless, as a robustness test, we run a Probit model with the same independent variables. The results are shown in Appendix 2.

We run two models. A linear OLS (1) and two model specifications with fixed effects for the executive tax auditor and the enterprise (organisation number) (2):

(1) $Y_i = \alpha + \beta_1 T_i + \beta_2 CPI_i + \beta_3 War_i + \beta_4 X'_i + \varepsilon_i$

(2) $Y_i = \gamma_1 T_i + \gamma_2 CPI_i + \gamma_3 War_i + \gamma_4 X'_i + \epsilon_i$

Both models have standard errors clustered by enterprise because we have sampled data from a population using clustered sampling (Liang & Zeger, 1993).

The dependent variable (Y) is the compliance of employee i. The variable takes a value of 1 if the employee holds a written contract and 0 otherwise.

CPI values are originally on a scale from 0-100 where countries with high corruption levels have low values. In our models, CPI_i is an inverted CPI score for the employee in the year of migration to Norway. The inversion is done to provide a more intuitive interpretation of coefficients, i.e., corruption levels increase with values. Norwegian employees have received a CPI score of value 16 for the audit year 2018.

War_i is an armed conflict dummy for employee i, which takes the value 1 if there has been a conflict in the last 25 years before migration to Norway, and 0 if not.

T is the employee's residence time in Norway as specified above. X is a vector of the following control variables for both models:

1) Gender dummy
2) Average, inverted CPI score for all employees in the same firm
3) Ratio of employees in the same firm with conflict background
4) Foreign dummy
5) Employee categories

Variables 2) and 3) are included to control for the effects of corruption and war in the enterprise at which the employee is hired. It may be that employees are influenced by colleagues' war or corruption backgrounds, not just their own. Furthermore, given that employees spend a notable number of hours at the workplace, we would assume that the work environment is a source of compliance influence from these factors as well.

We have also run models with interactions between residence time and inverted CPI score, and between residence time and conflict, to examine any interaction effects between the employee's background and residence time in Norway. We find no such effects. In other words, the corruption and conflict effects we find are independent of residence time.

4.4 Descriptive statistics

The main characteristics of the variables in the models are summarised in Table 1.

Table 1: Descriptive statistics

Variable	N	Average	SD	Min	Max
Written contract	4937	0.83	0.38	0.00	1.00
Residence time	5007	26.76	19.52	0.00	71.00
Female	5007	0.36	0.48	0.00	1.00
Conflict	5007	0.08	0.28	0.00	1.00
Conflict share Employee	5007	0.08	0.17	0.00	1.00
CPI Employee	5004	27.96	19.79	0.00	92.00
Average CPI Employee	5007	28.02	14.20	5.00	86.38
Foreign	5007	0.38	0.49	0.00	1.00
Skilled	4864	0.24	0.43	0.00	1.00
Manager	4864	0.16	0.37	0.00	1.00
Unskilled	4864	0.38	0.48	0.00	1.00
Other	4864	0.21	0.41	0.00	1.00

Note: Column 3 shows the mean value of the variable (Column 1) among employees in the sample. Column 4 shows the standard deviation. Written contract (dependent variable), Female, Conflict, Foreign, Skilled, Manager, Unskilled and Other are dummy variables. Female takes the value 1 if the employee is female. Residence time value is years. Conflict takes the value 1 if the employee was exposed to armed conflict in their home country up to 25 years prior to migration. CPI score is the inverted CPI score of the employee's country of origin at the year of arrival to Norway. Conflict share Employees is the fraction of employees in the firm exposed to armed conflict in the home country up to 25 years prior to migration.

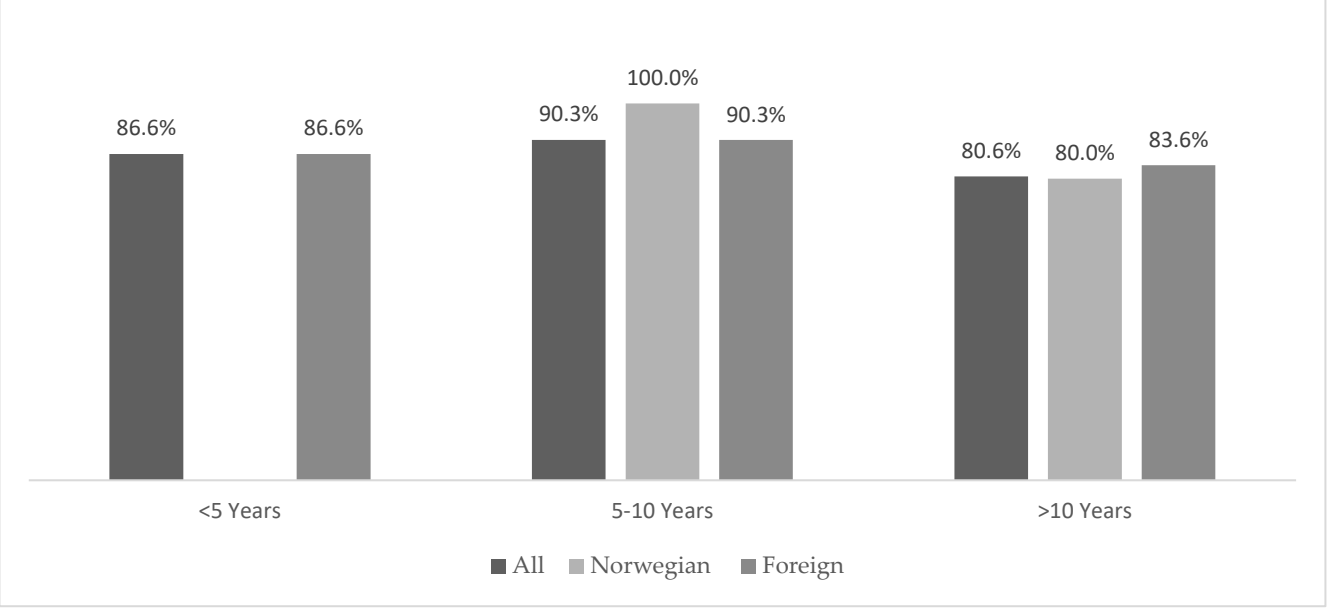
83 percent of employees have written contracts, which is our measurement of compliance. The average residence time is 26.8 years, average CPI score for the employees in the sample is 28. It is also the mean value when we look at the average CPI score for all employees in the enterprise. For other dummy variables, the "average" indicates the ratio of the variable taking the value of 1, that is 36 percent of the sample is female, 8 percent are exposed to conflict, 38 percent are foreigners, etc. Despite a relatively high number of foreign employees, we see that few come from a country undergone armed conflict less than 25 years before arrival in Norway.

Bastani et al. (2020) analyse differences in completed travel deductions in the Swedish tax return for the period 2002-2013 between Swedish-born and immigrants. In line with their study, we have also divided the population into residence groups (<5 years; 5-10 years; >10 years) based on the same reasoning; foreign workers become increasingly assimilated in the labour market over time, but there is no linear relationship between residence time and

claimed deductions. Nevertheless, the tendency is clear in their study; over time, foreign taxpayers claim more deductions and become more like the rest of the labour force.

If we break down compliance, that is, written employment contracts, into the different residence groups, we do not see a clear trend. It is not the case that compliance increases steadily with the number of years in Norway, see Figure 1.

Figure 1: Compliance and residence time in Norway (year)



Note: The first residence group consists entirely of foreign workers, and there are very few Norwegians in the second group. Norwegians in the second group are likely to be either foreign citizens who have acquired Norwegian citizenship, or Norwegians who have left the country for some time.

We see that the probability of holding a written employment contract increases from the first to the second residence group, and then decreases. A possible explanation for this may be that, after a certain period, one becomes more relaxed about formalising the employment relationship, that one "learns" the culture of informal agreements in these sectors, or that one gradually assumes that the probability for sanctions is lower than previously thought.

If we limit the sample to just foreign workers, it is only in the last residence group where we observe differences between foreign and Norwegian workers, simply due to the lack of Norwegian workers in the first groups. In this residence group, the foreign workers have a somewhat higher probability for written contracts than their Norwegian peers. This difference in compliance confirms the tendency in Bastani et al. (2020). In their study, however, the response variable is not compliance as such, but rather how much deduction is claimed in the tax return. The immigrants claim fewer deductions than their Swedish fellow citizens in the

first residence groups, and then claim more than the Swedes in the last group when they have lived longer in Sweden and presumably learned and adapted further to the tax system. In other words, there can be a learning mechanism behind both these results.

However, the tendencies we see for the share with contracts in the various residence groups may be due to other characteristics of the workers than residence time. The results become more interesting when we control for other, relevant characteristics, as well as characteristics that are constant across enterprises and responsible auditors in the NTA.

5. Results

5.1 Regression Results

The main findings from the regression models are shown in Table 2.

Table 2: Regression results.

Variable	(1)	(2)	(3)
Residence time <5 years	0.004	0.020	0.015
	(0.024)	(0.024)	(0.024)
Residence time 5-10 years	0.042*	0.054**	0.048**
	(0.023)	(0.024)	(0.024)
Conflict	-0.099***	-0.090***	-0.080***
	(0.030)	(0.029)	(0.030)
Foreign	0.072***	0.043*	0.046*
	(0.026)	(0.025)	(0.027)
Female	0.031**	0.025**	-0.022
	(0.013)	(0.013)	(0.014)
CPI Employees	0.000	0,001	0.000
	(0.001)	(0.001)	(0.001)
Average CPI Employees	-0.002*	-0.002**	
	(0.001)	(0.001)	
Conflict share Employees	0.139**	0.107	
	(0.067)	(0.067)	
Skilled	0.149***	0.149***	0.133***
	(0.020)	(0.020)	(0.025)
Manager	-0.028	-0.031	-0.046*
	(0.024)	(0.024)	(0.027)
Unskilled	0.077***	0.085***	0.109***
	(0.020)	(0.021)	(0.026)
Constant	0.759***		
	(0.027)		
N	4836	4794	4478
R^2	0.043	0.131	0.591
F	17.568	17.071	14.200

Note: Estimated coefficients from OLS model runs. Standard errors in parentheses. Column (1) is OLS without fixed effects, column (2) is OLS with fixed effects on the NTA auditor, and column (3) is OLS with fixed effects on Enterprise of the employee. Residence time >10 years is the reference category and therefore omitted in the table. Standard errors are clustered by Enterprise. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

5.2 Discussion of Findings

In the first model, compliance is significantly higher for those with a residence time between 5 and 10 years than for those with a residence time of more than 10 years (reference category). This result remains when we control for who has carried out the audit and the enterprise to

which the employee belongs. This pattern is similar to that of Bastani et al. (2020) who found that immigrants in the first residence time group file substantially less commuter deductions than natives, while immigrants in the second group behave more like natives, and, finally, immigrants with the longest residence time in Sweden file the most, even more than natives (Bastani et al., 2020). Thus, as Bastani et al. (2020) document a non-linear "deduction adaptation," we find a similar pattern of "compliance adaptation," suggesting that evolving norms erode the importance of tax honesty over time (Hofmann et al., 2017). The pattern of the residence time gradient makes it unlikely that age is the driver behind the coefficients. We have regressed compliance over the employee's age and found no significant effects.

Exposure to armed conflict before arrival in Norway results in a 9.9 percent reduced probability of compliance, but the effect decreases when we control for which auditor in the Tax Administration has executed the audit (9.0 percent), and the enterprise (8.0 percent). Ballentine & Sherman (2003) analyse the dynamics of war economies and point to mechanisms related to the expansion of the informal economy whereby the breakdown of formal institutions weakens enforcement. This mechanism may also be at play in this context.

We find no clear effect of corruption level, except for a marginal but statistically significant effect of the average corruption level of the employees in the enterprise. This variable is fixed for all employees in the same enterprise and is therefore not estimated for the last model specification that includes control for the enterprise. There is no statistically significant effect of employee CPI Score in any of the models.

One explanation may simply be that the workplace environment affects employee compliance behaviour far more than corruption exposure in the home country. The perception of strong and efficient enforcement in Norway can deter non-compliance regardless of corruption inheritance from the home country. In certain cultural or social contexts, there may be strong norms and values that influence tax compliance behaviour, and these norms can encourage individuals to fulfil their tax obligations based on a sense of civic duty, ethical considerations, or social pressure, overriding the negative influence of corruption inheritance.

Labour market benefits in Norway may create alternative incentives for compliance, as it may be necessary to hold a work permit and contract to qualify for government programs, receive loans or grants, or maintain a good reputation for business transactions. These alternative incentives can motivate tax compliance independently of corruption inheritance.

There is a significantly higher compliance rate for employees with foreign backgrounds, but this effect decreases when we control for the auditor and enterprise. The significance level concerning other effects is also somewhat decreasing when we control for the auditor and

enterprise. Female employees have a 3.1 percent higher probability of holding a written employment contract, but this effect disappears when we control for the enterprise.

The results suggest that the enterprise itself plays a role in shaping the compliance behaviour of employees with a foreign background, and the higher compliance rate among female employees.

A finding which does not remain independent of the auditor, however, is that the ratio of employees in the enterprise exposed to conflict results in a *higher* probability of compliance of the individual employee. When the ratio of employees with a conflict background increases by 10 percentage points, the probability of compliance *increases* by 1.39 percent. The result, which is statistically significant at 5 percent, appears counterintuitive compared to the opposite sign of the conflict dummy, expressing the employee's own conflict exposure.

If we compare these results, the situation appears such that the employee's conflict exposure negatively affects his or her compliance, but colleagues' conflict exposure positively affects his or her compliance. However, because the result is not sustained when we control for the responsible auditor, the effect is likely spurious.

If we look at the position categories, much of the variance in compliance can be explained by the type of position. Skilled workers are 14.9 percent more likely to hold a written employment contract than workers in other types of positions, even when we account for the auditor. The effect remains when we control for the enterprise but is somewhat weaker (13.3 percent). Correspondingly, unskilled workers are 8.5 percent more likely to hold a written employment contract than workers in other types of positions, even when we account for the auditor, and the effect is stronger when we control for the enterprise (10.9 percent). Managers are slightly less likely to hold a contract when we control for the business-specific characteristics.

It is not straightforward to identify potential mechanisms at play here. Managers often have greater autonomy and decision-making authority in their organizations. This may result in less direct oversight of their obligations by superiors or a lack of strict reporting mechanisms. In contrast, skilled and unskilled employees may be subject to more scrutiny and oversight from their employers or supervisors, leading to higher compliance rates. Skilled and unskilled employees may perceive holding a written contract as crucial for maintaining a positive professional reputation and securing future career opportunities. Non-compliance can have severe consequences for their employment prospects or professional growth. On the other hand, managers may have more leverage or resources to handle potential tax issues, which could influence their compliance behaviour. Skilled and unskilled employees often work closely with others and may be more socially connected within their organizations or communities. The visibility of their compliance behaviour and the potential for peer pressure

or social expectations may contribute to higher compliance rates. Managers, depending on their level of interaction with others or their position within the organization, may face different social dynamics that may influence their compliance behaviour.

Holding a formal employment contract is likely to be correlated with other employee characteristics as well. Type and level of education, willingness to work, and language training are a few examples that one would expect will influence compliance. Educational factors are probably positively correlated with an employee's knowledge of rights and duties, including knowledge about holding an employment contract. Hence, unobserved variables may confound the associations we observe if they are related to both compliance behaviour and our key independent variables. The Tax Administration has no information on e.g., the employee's field or level of education, or willingness to work, and so omitted variable bias is likely to compromise the results.

Different characteristics and compliance attitudes may, however, be selected to different firms, in that employees with a high education may be attracted to the same firm. We thus include fixed effects on enterprise. With this approach, we compare the compliance level among foreign and Norwegian employees working at the same company. With fixed effects on enterprise, we find that the significance level and coefficients are by and large unaffected.

There are reasons to believe there is bias in the data consistent with endogeneity. Unobserved factors, such as personal motivation, language skills, or cultural integration, could contribute to the observed associations. It is possible that employees who have e.g., a higher likelihood of staying in Norway or acquiring Norwegian citizenship are more likely to secure a formal employment contract. In other words, the causality can be reversed, with the longer residence time or intention to acquire citizenship leading to a higher probability of obtaining formal employment. There may also be underlying factors that influence both the likelihood of having a formal employment contract and the residence time and acquiring Norwegian citizenship.

There may also be endogeneity between conflict exposure and compliance through the asylum institute. Factors such as socioeconomic conditions, political instability, educational opportunities, or personal networks can impact both migration decisions and the ability to obtain asylum.

5.3 Robustness

The random audits in the Tax Administration are based on a template consisting of fixed audit actions, unlike risk-based audits, which include more case-specific, investigative actions. When we account for auditor, statistical significance decreases in the OLS models for variables, foreign background and gender, and disappears completely for the enterprise conflict ratio. This may be due to the systemic differences in the assignment of audit subjects. Although the

enterprises are randomly audited, they are not randomly assigned to auditors. For example, it is conceivable that more experienced auditors have been assigned enterprises that appear more demanding to audit or enterprises that are assumed to have a higher risk of non-compliance.

A high correlation between different independent variables gives reason to test for multicollinearity. When there is a strong linear relationship between the independent variables, it is not given that the coefficients of a regression modal can be uniquely calculated. As the degree of the multicollinearity increases, the estimated coefficients become unstable and their standard errors can be overstated (Chen et al., 2005). One way to measure multicollinearity is to look at the so-called variance inflation factor (VIF) of the individual variable, where levels higher than 10 reveal the need for other robustness tests. However, none of the variables have a higher VIF than 6, and we can therefore disregard multicollinearity.

In econometric regression analyses, several hypothesis tests are often conducted simultaneously, and this also applies to the analyses in this article. The problem then becomes how to decide which hypotheses can be rejected, or more precisely, whether significant effects that appear after many, different hypothesis tests are real or spurious. Romano and Wolf (2005) suggest a stepwise test procedure that compared to related test methods is "more powerful" and will more often reject false hypotheses. Unlike some stepwise methods, Roman-Wolf implicitly captures common dependence structures in the test statistics, something that results in an increased ability to uncover false hypotheses. Our assessment is therefore that it provides even more robust tests than more "traditional" tests for multiple hypotheses such as Benjamini and Hochberg (1995) or Bonferroni (1936).

We have run tests à la Romano-Wolf on all the model variants and find that all corrected p-values resulting from the test are robust, cf. Appendix 7.

We have also compared the results from the first simple OLS model without controlling for the auditor or enterprise with results from a Probit model according to the same specification. The results shown in Appendix 7.2 reflects only minor deviance between Probit and OLS.

6. Conclusion

There are several reasons why compliance among foreign workers is of particular interest to the NTA. Norway has high labour immigration relative to the population, employment is highest in labour-intensive industries with a large ratio of unskilled labour, and the industries have been subject to a large allocation of the administration's audit resources. What affects compliance among foreign workers in these sectors is, as for taxpayers in general, complicated.

However, for these workers, certain cultural differences can affect the degree of compliance, and we assume that these become smaller over time. Therefore, we have examined whether

residence time in Norway affects compliance, given the level of corruption and exposure to armed conflict in the country of origin before migration to Norway.

Generally, we find no statistically significant importance of residence time in Norway for these workers, except an increased compliance among those with residence time of between 5 and 10 years compared to those with residence time over 10 years, when we look at all workers together. The effect strengthens, and uncertainty in the estimates decreases when we control for the auditor and enterprise. We find no significant effect on the individual corruption level, but a statistically significant reduced probability of compliance, if the worker has been exposed to armed conflict up to 25 years before arrival in Norway, which also is sustained when we control for the auditor and enterprise.

7. Appendix

7.1 Target population, sample and weights

Table A1 provides an overview of the number of registered employment relationships in the target population and the sample weight that indicates whether the corresponding industry (stratum) is over- or under-represented in the sample. We see that the weights for employees deviate from the weights for enterprises. This is because a sample of employees has been drawn from each of the enterprises within a stratum, which is not proportional to the sample of enterprises in the same stratum.

In addition, the number of the target population is calculated based on the 1,974 enterprises and not based on the 1,901 enterprises that are represented in the sample of registered employment relationships. The reason we use the 1,974 enterprises as a basis is that we do not have a full overview of the reason why the 73 enterprises that are dropped are not represented. However, we have corrected the number of employees in the target population in proportion to the reduction we made in the number of enterprises in the target population. In other words, the number of registered employment relationships in the target population has been corrected for 72 enterprises that had ceased or, for some other reason, could not be reached prior to the audit activities.

Table A1: Registered employment relationship: Target population, sample and weights

Industry	(1)	(2)	(3)
No Nace code	2,042	83	0.525
Other land transport with passengers (taxi, tour bus, etc.)	18,088	106	3.642
Mining and quarrying	1,335	67	0.425
Retail trade, except with automobiles	4,826	89	1.157
Fisheries, catches and aquaculture	6,226	122	1.089
Business services, except cleaning businesses	24,904	530	1.003
Hairdressing, beauty care and body care	11,177	205	1.164
Health and social welfare services	31,085	226	2.936
Industry	27,906	352	1.692
Information and communication	30,909	275	2.399
Agriculture and services related to agriculture, hunting and wildlife care	4,362	427	0.218
Cultural enterprises, entertainment and leisure activities	9,935	183	1.159
General land transport (mainly freight transport)	18,413	316	1.244
Painting, flooring (other finishes without carpenter)	8,079	376	0.459
Construction of buildings and construction activity	47,756	1,046	0.974
Accommodation	8,962	387	0.494
Cleaning enterprises	8,685	587	0.316
Catering enterprises	59439	1,337	0.949
Carpentry work	5,704	298	0.409
Transport and storage excluding land transport	10,022	172	1.244
Water supply, sewage and garbage enterprises	1,905	64	0.635
Maintenance and repair of motor vehicles, except motorcycles	13,376	332	0.860
Total	355,135	7,580	

Note: Column (1) is the population number of registered employment relationships, column (2) is the sample number of registered employment relationships in sample, and column (3) is the weights (over-/under-representation).

7.2 Results from OLS and Probit Model Comparison

We have run an OLS model (1) and a Probit model (2):

(1) $Y_i = \alpha + \beta_1 X'_i + \varepsilon_i$

(2) $\Pr(Y_i = 1|X) = \Phi(\beta_1 X'_i) + \varepsilon_i$

The OLS model is identical to our main model, but for the sake of simplicity, **X** now represents a vector of *all* independent variables. In the Probit model, the dependent variable is the probability of compliance (written employment contract), given the same vector of independent variables. Φ is the cumulative distribution function (CDF) of the standard normal distribution.

In the table below, we have compared marginal effects (probability of compliance, given that other variables in the model are set to average) for the variables that make this comparison possible and meaningful across the two model specifications. As we can see, the two model specifications show only slight differences when we compare the marginal effects of these variables.

Table A2: OLS and Probit marginal effects comparison

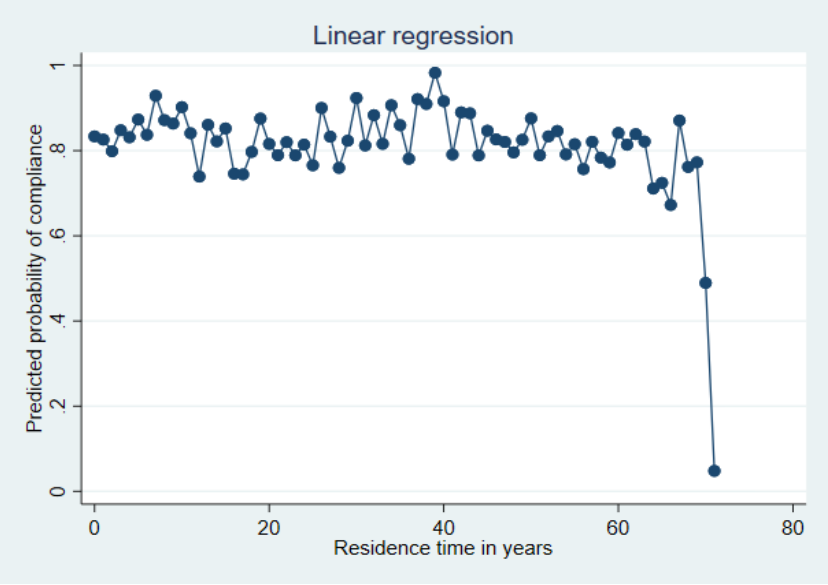
Variable	(1)	(2)
Residence time <5 years	82.8%	82.8%
Residence time 5-10 years	86.6%	87.4%
Residence time >10 years	82.4%	82.4%
Conflict	73.8%	70.9%
Foreign	87.3%	87.4%
Position category: Skilled	91.8%	91.5%
Position category: Manager	74.0%	74.7%
Position category: Unskilled	84.5%	84.6%
Position category: Other	76.8%	76.8%

Note: Column (1) shows predictive margins from OLS model and column (2) shows predictive margins for the Probit model.

7.3 Compliance and Residence Time

We have regressed compliance over all residence year dummies and plotted the coefficients. There appears to be no linear relationship between residence time and predicted probability of compliance. Compliance is low among those with the longest residence time, but this particular dip is driven by a small number of employees without employment contracts in this part of the sample.

Figure A3: Predicted compliance over residence year dummies



7.4 Romano-Wolf Test

Table A3: P-values Test Statistics

Variable	(1)	(2)	(3)
Residence time	0.000	0.010	0.010
Conflict	0.613	0.654	0.654
Foreign	0.000	0.010	0.010
Female	0.008	0.020	0.020
CPI Employees	0.000	0.010	0.010
Conflict share Employees	0.228	0.267	0.267

Note: P-test statistics are given by the original p-values (1), resample p-values (2) and Romano Wolf p-values (3) of each independent variable.

8. References

Abraham, M., Lorek, K., Richter, F., & Wrede, M. (2015). *Collusive Tax Evasion and Social Norms*. https://EconPapers.repec.org/RePEc:ces:ceswps:_5167

Alesina, A., & Angeletos, G.-M. (2005). Corruption, inequality, and fairness. *Journal of Monetary Economics*, *52*(7), 1227-1244.

Ali, M., Fjeldstad, O.-H., & Sjursen, I. H. (2013). Factors affecting tax compliant attitude in Africa: evidence from Kenya, Tanzania, Uganda and South Africa. Centre for the Study of African Economies 2013 Conference, Oxford University,

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A Theoretical Analysis. *Journal of Public Economics*, *1*(3-4), 323-338.

Alm, J. (2019). What motivates tax compliance? *Journal of economic surveys*, *33*(2), 353-388.

Azam, J.-P., & Mesnard, A. (2003). Civil war and the social contract. *Public Choice*, *115*(3-4), 455-475.

Ballentine, K., & Sherman, J. (2003). *The political economy of armed conflict: Beyond greed and grievance*. Lynne Rienner Publishers.

Bastani, S., Giebe, T., & Miao, C. (2020). Ethnicity and tax filing behaviour. *Journal of Urban Economics*, *116*, 103215.

Baum, A., & Gupta, S. (2017). Corruption, taxes and compliance. *eJTR*, *15*, 190.

Bellows, J., & Miguel, E. (2006). War and institutions: New evidence from Sierra Leone. *American Economic Review*, *96*(2), 394-399.

Benjamini, Y., & Hochberg, Y. (1995). 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.

Bjørneby, M., Alstadsæter, A., & Telle, K. (2018). Collusive Tax Evasion by Employers and Employees: Evidence from a Randomized Field Experiment in Norway.

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, *8*, 3-62.

Brautigam, D., Fjeldstad, O.-H., & Moore, M. (2008). *Taxation and state-building in developing countries: Capacity and consent*. Cambridge University Press.

Caudill, S. B. (1988). An advantage of the linear probability model over probit or logit. *Oxford Bulletin of Economics and Statistics*, *50*(4), 425-427.

Chen, X., Ender, P. B., Mitchell, M., & Wells, C. (2005). *Stata web books: regression with Stata*. https://stats.idre.ucla.edu/stata/webbooks/reg/stata-web-books-regression-with-stata/

Clausen, B., Kraay, A., & Nyiri, Z. (2011). Corruption and confidence in public institutions: Evidence from a global survey. *The World Bank Economic Review*, *25*(2), 212-249.

Collier, P., & Sambanis, N. (2002). Understanding civil war: A new agenda. *Journal of Conflict Resolution*, *46*(1), 3-12.

Damania, R., Fredriksson, P. G., & Mani, M. (2004). The persistence of corruption and regulatory compliance failures: theory and evidence. *Public choice*, *121*(3-4), 363-390.

Eriksen, K., & Fallan, L. (1996). Tax knowledge and attitudes towards taxation; A report on a quasi-experiment. *Journal of Economic Psychology*, *17*(3), 387-402.
Galletta, S., & Giommoni, T. (2023). War Violence Exposure and Tax Compliance. *Available at SSRN 4316553*.
Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M., & Strand, H. (2002). Armed conflict 1946-2001: A new dataset. *Journal of peace research*, *39*(5), 615-637.
Harding, L., & Mutascu, M. (2016). *Does migration affect tax revenue in Europe?*
Hofmann, E., Voracek, M., Bock, C., & Kirchler, E. (2017). Tax compliance across sociodemographic categories: Meta-analyses of survey studies in 111 countries. *Journal of Economic Psychology*, *62*, 63-71.
Holte, H. C. (2020). Å bygge tillit er viktig, men kontroll trengs. *Dagens Perspektiv*. https://www.dagensperspektiv.no/2020/a-bygge-tillit-er-viktig-men-kontroll-trengs
Jahnke, B., & Weisser, R. A. (2019). How does petty corruption affect tax morale in Sub-Saharan Africa? *European Journal of Political Economy*, *60*, 101751.
Kirchler, E. (2007). *The economic psychology of tax behaviour*. Cambridge University Press.
Kleven, H. J., Landais, C., Saez, E., & Schultz, E. (2014). Migration and wage effects of taxing top earners: Evidence from the foreigners' tax scheme in Denmark. *The Quarterly Journal of Economics*, *129*(1), 333-378.
Kolm, A.-S., & Larsen, B. (2019). Underground activities and labour market performance. *International Tax and Public Finance*, *26*(1), 41-70.
Lambsdorff, J. G. (1999). Corruption in empirical research: A review. *Transparency International, processed*, *6*.
Lassen, D. D. (2007). Ethnic divisions, trust, and the size of the informal sector. *Journal of Economic Behaviour & Organization*, *63*(3), 423-438.
Leitao, A. (2016). Corruption and the Environment. *Journal of Socioeconomics*, *5*(3).
Liang, K.-Y., & Zeger, S. L. (1993). Regression analysis for correlated data. *Annual review of public health*, *14*(1), 43-68.
Litina, A., & Palivos, T. (2016). Corruption, tax evasion and social values. *Journal of Economic Behaviour & Organization*, *124*, 164-177.
Medina, L., & Schneider, F. (2017). Shadow economies around the world: New results for 158 countries over 1991-2015.
Miguel, E., Saiegh, S. M., & Satyanath, S. (2011). Civil war exposure and violence. *Economics & Politics*, *23*(1), 59-73.
Nordblom, K., & Žamac, J. (2012). Endogenous Norm Formation Over the Life Cycle-The Case of Tax Morale. *Economic Analysis & Policy*, *42*(2).
Nygard, O. E., Slemrod, J. B., & Thoresen, T. O. (2016). Distributional Implications of Joint Tax Evasion.
Pissarides, C. A., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, *39*(1), 17-32.
Riley, M. W., & Riley Jr, J. W. (1989). The lives of older people and changing social roles. *The Annals of the American Academy of Political and Social Science*, *503*(1), 14-28.

Romano, J. P., & Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, *100*(469), 94-108.
Seleim, A., & Bontis, N. (2009). The relationship between culture and corruption: a cross-national study. *Journal of Intellectual Capital*, *10*(1), 165-184. https://doi.org/10.1108/14691930910922978
Slemrod, J. (2016). Tax compliance and enforcement: New research and its policy implications.
Slemrod, J. J. J. o. E. P. (2007). Cheating ourselves: The economics of tax evasion. *21*(1), 25-48.
Torgler, B. (2014). *Can tax compliance research profit from biology?*
Transparency International. (2019a). *Corruption Perceptions Index*
Transparency International. (2019b). *Corruption Perceptions Index 2018: Score timeseries since 2012*, . https://www.transparency.org/en/cpi#
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
Young, C., & Varner, C. (2011). Millionaire migration and state taxation of top incomes: Evidence from a natural experiment. *National Tax Journal*, *64*(2), 255.

Tax Compliance among Managers. Evidence from Randomized Audits[4]

Thomas Lange[5]

Anne May Melsom[6]

Abstract

Earlier studies on individuals' law-abiding behaviour have found significant effects of their home country's corruption level on compliance. In our study of managers' tax compliance, we use data from random audits and find associations between tax compliance and the use of an external accountant, age, managers' origin and employees' conflict exposure, but no effect of managers' conflict exposure, nor CPI Scores. The use of an external accountant seems to commit managers to comply with reporting requirements. Our findings suggest that factors such as managers' age, and company characteristics are important in understanding manager compliance. To study whether other mechanisms not previously tested may provide explanatory power, we specify two Machine Learning (ML) models, which confirm our findings and suggest other associations.

JEL Classification: H2; H26; M41; M48; M54

[4] Special thanks to Nils Gaute Voll for guidance on Machine Learning models. Thanks to Evelina Gavrilova-Zoutman and Floris Zoutman for rigorous comments on various topics in this paper.

[5] Corresponding author, Norwegian School of Economics.

[6] Norwegian Tax Administration

1. Introduction

Norway has one of the world's most generous welfare states, primarily financed by tax revenues. Since the expansion of the EU in 2004, Norway has seen a significant increase in the number and mobility of foreigners in the workforce. With a relatively stable workforce concerning the diversity of countries represented post World War II, Norway has experienced a large shift in this composition in the past 20 years. In 1990, approximately 3.5 percent of the Norwegian population was foreign-born, compared to 14.8 in 2021 (Statistics Norway, 2022). Since the year 2000, a large number of immigrants have come from countries with different taxpayer cultures.

Obtaining a high level of tax compliance is essential for tax administrations worldwide, and managers play an important role in setting the standards for adequate compliance levels. Managers are liable to several additional reporting requirements, and so their compliance behaviour may be affected by other factors than their individual compliance preferences. What drives managers' tax compliance may be determined by both intrinsic factors, such as cultural and social norms, and factors related to their tax reporting duties. Immigrants are of particular interest to study in this context because they typically bring cultural and social norms from their country of origin into a new social and cultural environment in their destination country (Foner (2014); Potocky and Naseh (2020)). Such effects are of rising importance to modern tax administrations in the era of globalization and international exchange of the workforce.

Whereas the literature by and large focuses on individual taxpayer compliance, the drivers behind tax compliance among managers, are less studied. In a third-party reporting regime of modern tax jurisdictions, the need for a better understanding of manager compliance is important because pre-filled individual tax declarations have enabled the tax administration to lift the burden of upholding its trust from individual taxpayers to managers.

In this paper, we use data from random audits carried out by the Norwegian Tax Administration (NTA) to ask which factors drive compliance among firm managers in labour-intensive sectors. The purpose is two-fold. First, we use an OLS fixed effects model to estimate the marginal effects of how characteristics of managers, known from previous literature, affect the likelihood of compliance. Second, we introduce two ML models, namely one Genetic Algorithm (GA) and one Least Absolute Shrinkage and Selection Operator (LASSO) to perform

variable selection of the full data set, which allows us to both confirm the findings from the OLS model and study a much wider array of factors associated with manager compliance.

While we specify the OLS on individual characteristics of the manager, the variable selection from the ML models points towards characteristics of the firm. The way our ML models are specified, they do a good job in selecting control variables we estimate with OLS as well, and compare the results with our 'standard' OLS model. This twofold strategy allows us both to test 'common' independent variables from other compliance areas, and to systematically explore a magnitude of reporting variables not previously tested scientifically in the literature. In this context, we do not use the ML models for prediction, but to perform variable selection, *even though* the results could be interpreted as predictions (Battaglini et al., 2022).

The randomized audits of Norwegian companies concern wage and labour regulation compliance. The on-site audits were executed on a stratified random sample of 1,974 firms in labour-intensive businesses in Norway during the filing and auditing season of 2018. The audits included many questions, checkpoints and control actions covering several reporting liabilities. Audited firms were notified shortly prior to the audit, about which information the NTA would collect. The empirical analysis is divided into two main parts.

The first part studies how Managers' cultural background and residence time is associated with compliance using audit data combined with the Corruption Perception Index (CPI) (Transparency International, 2019b) of birth country and the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al., 2002) as institutional proxies for taxpayer culture.

The second part of our analysis is a suggested solution to the problem of variable selection, where we allow for non-parametric inference. We investigate 100 prospective independent variables suggested by NTA auditors contained in our data set. Unlike previous work based on random audit samples, such as Kleven et al. (2011), we extend our analysis by using two different supervised algorithms to perform variable selection from our sample.

Our approach contributes to the current research field by using a wide array of randomized audit data, which enhances the generalizability of findings. Our study is more representative of real-world tax compliance scenarios among managers, making the results applicable to a broader range of contexts. The data from Managers' 'natural' tax environment allows for the

identifying patterns and trends that may not be apparent in smaller or more narrowly focused studies. By utilising ML models from other research fields, we unravel a more complete understanding of tax compliance among managers. Both LASSO and GA are known for their variable selection capability from other fields of research, which helps in identifying the most relevant features influencing tax compliance, given both the complexity and quality of our data set. As such, the application of advanced machine learning models introduces methodological diversity to tax compliance research. Using well-defined and tested algorithms, we broaden the methodological toolkit available for studying tax compliance, enabling researchers to develop new approaches to the research field.

In our OLS model framework, we find that employees' previous exposure to armed conflict is associated with less likelihood of compliance, but no such associations between managers' own conflict exposure and compliance. We find positive associations between compliance and age, and the use of an external accountant robust across all model specifications. However, we find no significant effects of home country corruption level. These results are consistent with Kleven et al. (2011) who find that the impact of social and cultural variables is small compared to variables that capture information and incentives. The result also indicates that there are other factors driving tax compliance among managers compared to individual taxpayers. The larger number of reporting liabilities, and thus contact points with the tax administration, may increase both the learning effect from frequent reporting, and the perception of tax authority presence and enforcement.

We find that managers of private limited companies (hereafter Private Ltd) are significantly more likely to be compliant than managers of other organisational forms. GA and LASSO reproduce the positive effects from the OLS models of holding an external accountant. Furthermore, holding a salary system and the number of terms the firm reports salaries or social benefits for the employees have both a positive impact on managers' compliance. We find a negative effect of foreign employees' exposure to armed conflict. One explanation could be that the potential stress, trauma, or altered perspectives resulting from conflict experiences among employees may indirectly affect the decision-making within the managerial hierarchy. These findings are robust across all 3 models.

By providing insights into the key determinants of tax compliance among managers, our findings can inform policymaking and strategies for tax administration. Understanding which

factors significantly influence compliance allows policymakers to develop more targeted and effective interventions.

The paper is organized into five sections. Section 2 gives a short review of the relevant literature, and how this paper fills the gaps in the existing body of empirical work studying tax compliance. Section 3 presents the data. Section 4 describes the three models, namely the OLS, the GA and, finally the LASSO. Section 5 estimates the effects of cultural background and residence time (OLS), defines two models from variable selection performed by the GA and LASSO, respectively, and then estimates the effects from these. Section 6 concludes.

2. Related literature

Since Torgler (2007) and Torgler et al. (2008) studied cross-cultural differences in tax compliance behaviour, research on tax morale is typically concerned with both the underlying morale as a cause of tax evasion (Cummings et al., 2009), and, slightly contrary, why tax morale in Western countries are so high, given the low probability of audits and sanctions (Frey & Torgler, 2007). While there is some evidence that increasing ethnic fractionalization decreases voluntary tax compliance (Lassen, 2007), the historically homogeneous population in e.g. Norway has traditionally left a whole generation of taxpayers with trusted responsibility to declare their income truthfully.

Previous literature suggests several 'cultural' mechanisms which may come into play when studying the associations between culture and tax compliance. Individualistic cultures tend to emphasize personal freedom, autonomy, and self-interest, while collectivist cultures prioritize social cohesion, harmony, and group interests. The review by Marandu et al. (2015) suggests that individualistic cultures may exhibit lower tax compliance due to a greater focus on self-interest and a weaker sense of duty towards the broader society. Trust and social capital may also influence tax compliance. Societies with higher levels of trust and strong social networks may exhibit higher tax compliance as individuals are more likely to cooperate and comply with tax obligations due to social norms and expectations. Gangl et al. (2016) give some support for this claim, that social capital fosters important prosocial behaviour and increases citizens' cooperation with the state. Bornman (2015) finds that positive attitudes towards government and perceptions of government legitimacy is likely to increase tax compliance. Conversely, if individuals perceive the government as corrupt, ineffective, or illegitimate, they may be less inclined to voluntarily comply with tax laws (Torgler, 2004).

Societies with a stronger emphasis on fairness and support for wealth redistribution and social justice may exhibit higher tax compliance rates (Hofmann et al., 2008). Bobek et al. (2013) find that social norms have direct as well as indirect influences on tax compliance behaviour. Societies with strong norms of honesty, fairness, and cooperation may thus exhibit higher levels of tax compliance due to the influence of social pressure and reputational concerns.

To our knowledge, there are few previous studies of drivers behind poor tax compliance among managers except for Joulfaian (2000) which attributes manager tax non-compliance to understatements in their own personal income tax, and some developing countries perspectives on SMEs under completely different tax liability regime, with little or no third-party reporting (Musimenta et al. (2017); Atawodi and Ojeka (2012). While Joulfaian (2000) finds that noncompliant firms are more likely to be managed by executives understating their personal taxes, the paper suggests few other controls that may form managers' preference for non-compliance.

Alm et al. (1992), Andreoni et al. (1998), Feld and Frey (2002), and Feld and Frey (2007) argue that factors such as social norms, tax morale, patriotism, guilt and shame explain variance in tax compliance. Tsakumis et al. (2007) find that national culture, as proposed by Hofstede (1984), is a significant factor in explaining tax evasion levels across 50 countries. Thus, depending on the options for non-compliance, e.g., the level of third-party reporting, taxpayers may be affected by non-economic drivers in their compliance behaviour.

While studies have found tax morale in the country of origin to be a significant determinant of tax morale among immigrants in the destination country (Kountouris and Remoundou (2013), few attempts have been made to test the common assumption in various tax administrations that immigrants become more compliant with time spent in the destination country.

The study of corruption norms and tax morale has gained increased attention, but the field is still small. Alm et al. (2016) and Jahnke and Weisser (2019) find that corruption drives higher levels of evasion. But their focus is the evasive effects of bribery of tax officials in the home country context, rather than the corruption level of the environment in which these tax officials operate. There are some examples of similar focus to our paper in the literature, however. Cummings et al. (2009) found a significant correlation between tax compliance behaviour and

tax morale in South Africa and Botswana, using CPI scores as one proxy for tax morale. Fisman and Miguel (2007) use another index for corruption, namely that of Kaufmann et al. (2005), and find a strong effect of corruption norms on diplomat parking violations, and a significant effect on enforcement. However, their environment deviates from ours in several respects. They study less complex parking regulations whereas our focus is complex tax reporting liabilities. In their context, diplomats have few contact points with legal authorities whereas managers in our context have more frequent contact points with the NTA. Parking regulations are not part of diplomats' day-to-day business, whereas reporting obligations are integrated into the managers' responsibilities running a legal business.

DeBacker et al. (2015) find that corporations with owners from more corrupt countries evade more US tax than owners from less corrupt ones. Common for earlier approaches is either the use of survey data or the use of risk-based tax audit data. An exception is Bastani et al. (2020) who utilise population-wide register data to investigate differences in the use of commuter deduction in tax filing between immigrants and Swedish natives. They find less filing among recently arrived immigrants than their native fellows. But their study concerns neither manager tax compliance nor their reporting liabilities, but rather filing of legitimate claims or lack thereof. However, the time effects they describe may be relevant for our study as well, since residence time in Norway also may affect managers' compliance.

Alm (2019b), Bjørneby et al. (2018), Kleven et al. (2011), and Alm et al. (2006) find that third-party reporting has increased compliance among individual taxpayers, simply because the opportunities for evasion have been reduced. Whenever third-party reporting is low, Kleven et al. (2011) show Allingham and Sandmo (1972) still has merit; when the options for evasion are present, evaded tax is a function of the probability of detection and a penalty for withholding. Some have shown that various tax administration measures, such as shifting tax remittance (Kopczuk et al., 2016) and public disclosure (Bø et al., 2015) lead to improvements in tax compliance. While the opportunities for evasion diminish with increasing levels of third-party reporting, this may not be the case for managers.

Shifts in tax compliance due to temporary or permanent demographic changes in the taxpayer population have, to our knowledge, not been under scrutiny. There may be several reasons for this. Rapid demographic changes in a population are often caused by immigration. Some studies exist on the effects of tax revenue from migration, but these few focus either on macro

revenue effects of immigration (e.g. Harding and Mutascu (2016), differences in the use of tax deduction among various immigrant groups in Sweden (Bastani et al., 2020), or the effects of tax rates on the migration of top income earners (e.g. Young and Varner (2011); Kleven et al. (2014)). Neither discuss compliance-related effects from migration or demographic changes. In other words, in the economic literature, there are few contributions on the connection between migration, demographic changes and tax compliance, and especially at the management level.

ML and GA in economics have previously been applied in stock market price predictions (Sable et al., 2017), financial asset portfolio selection (Sefiane & Benbouziane, 2012) or determination of real estate prices (Del Giudice et al., 2017). Such models have various applications in variable selection problems (Broadhurst et al. (1997); Tolvi (2004); Cateni et al. (2010); Liu and Ong (2008)), but to our knowledge, GA models are not used in the field of taxation. A more commonly used ML algorithm in econometrics is the LASSO (Hansen and Liao (2019); Fonti and Belitser (2017); Mullainathan and Spiess (2017); Belloni et al. (2012); Pereira et al. (2016); Chalfin et al. (2016)).

3. Data

We use data from randomized audits of Norwegian companies on wage and labour regulation compliance. Auditors collected company information and documentation on 1,974 random firms in labour-intensive businesses in Norway. The audits covered many reporting liabilities, primarily within the area of wage reporting, e.g., salary accounts, documented time use, tax deductions, staff registers, employment contracts, timesheets, overtime payments etc. Interviews and meetings with NTA auditors in the early stage of the random audit program resulted in a comprehensive list of 303 variables, of which 100 could be used for regression analysis.[7] The audits in our study focus on business reporting routines and standards rather than undeclared income. Thus, we use correct tax deductions, the existence of payroll accounts, correct monthly reporting, and the existence of general accounts to define compliance. The employer/manager bears the full responsibility to ensure compliance in all areas covered by our dependent variable.

Combining the data with information on residence time in Norway for the manager (owner or CEO) of the firm, and the Corruption Perception Index (CPI) of the birth country, the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al., 2002) as institutional proxies for

[7] Free text variables and variables of interest only for NTA internal administrative purposes were excluded.

taxpayer culture, we can study the influence of residence time on tax compliance among managers. The audited companies have been randomly selected, but they were not randomly assigned to auditors. Thus, we use auditor fixed effects to control for systematic differences between auditors, and fixed effects on Nace code (2-digit) to control for systematic differences between sectors.

3.1 Population and Sample

The 1,974 audited companies are randomly drawn from a target population of about 31,000 companies, representing 13 percent of all Norwegian businesses. When defining this target population, the NTA started with all companies defined as legal employers who reported working conditions and other tax-relevant information in the monthly reporting (A-melding) to the NTA during the year 2017 (Norwegian Tax Administration, 2019). The NTA then restricted this main population to target the audits to industries and businesses more at risk of non-compliance. Hence, the randomization is not valid for the whole taxpayer population, but rather a sub-population where the risk perception is higher than average. This limitation is based on a total of 14 criteria which are explained in more detail in Appendix Table A1a. It implies that the number of industries was reduced from 38 to 22 and that the number of businesses was reduced from 231,753 to 30,961.

Country of birth and citizenship are taken from the National Population Register. Information about the company's revenues and expenses in 2016 is taken from business reports to the Tax Administration (Income Statement I and II), while turnover is taken from the VAT Register. Information on NACE codes, establishment dates and termination dates are obtained from the Register of Legal Entities.

The target population is stratified according to the industry branch specified in the Appendix, i.e., each industry branch represents a stratum, or industry sector. The purpose of the stratification is primarily to enable a more effective selection for the audits, and to draw more businesses from industries with a large proportion of working conditions associated with increased risk of non-compliance. The sample selection method is proportional to the number of reported, foreign employees who arrived in the last three years, combined with a minimum and maximum number of businesses in the lower and upper part of the distribution.

3.2 Variables

The main dependent variable is Firm Compliance, which is an individual score of the most central, responsible person in the firm, typically a CEO or an owner. The dependent variable is continuous in reported income or tax liability as in e.g. Kleven et al. (2011), but rather a binary variable intended to measure compliance by any individual manager responsible for the firm's reporting liabilities, similar to the approach of Fellner et al. (2013). It takes the value 1 if there is no error in the 4 variables for the year 2017: tax deduction, existence of payroll accounts, correct monthly reporting and existence of general accounts.

Managers are liable to more reporting and contact points with the tax administration, and thus more subject to regulatory oversight. Even if the perceived audit probability is low, the frequency of contact points (e.g., the monthly reporting) may create a stronger incentive for managers to comply, although the evidence on this effect seems to go in both directions (Snow & Warren, 2005). Furthermore, managers are often responsible for establishing and implementing internal governance mechanisms within the firm. They have the authority to develop and enforce policies and procedures to ensure compliance with various regulations, including tax liabilities. Failure to enforce tax compliance within the organization can be seen as a failure of managerial responsibility, leading to potential internal control and governance issues (Alm, 2019a). Managers, therefore, have a vested interest in promoting tax compliance as part of their overall responsibilities. Finally, managers with ownership or financial holdings in the firm, have a direct financial interest in maintaining compliance. Tax compliance ensures the stability and sustainability of the firm's operations (Bird & Davis-Nozemack, 2018), but there is mixed evidence for a positive association between tax compliance and a firm's financial performance and value (Watson, 2015).

In our context, manager compliance is closely related to reporting obligations where only the manager is liable, not the employee. Thus, there are different evasion schemes available to the manager than to an employee, and thus more opportunities to evade. However, to which extent noncompliance can be due to honest mistakes or deliberate withholding of information, is still difficult to assess. The random audits focus on business reporting routines and standards rather than undeclared income. Tax deduction is one direct route in which the manager may underreport. Not having payroll accounts and general accounts of the firm, represent two other opportunities for non-compliance. The firm may intentionally underreport the number of employees or misclassify them as independent contractors or consultants. By

doing so, they can avoid the obligation of deducting and remitting payroll taxes. The firm also may engage in cash transactions or pay employees "off the books" without proper documentation or recording. This allows them to avoid reporting the income and associated payroll taxes.

An error in the firm's monthly reporting can involve an employer deliberately reporting lower income for employees than what is actually paid. This can lead to employees evading taxes on unreported income, but it presumes some sort of collusion between the employer and employee, especially in a tax regime with high levels of third-party reporting, cf. Bjørneby et al. (2021). Other items in the monthly reporting may also provide opportunities for evasion, like providing incorrect information about deductible expenses or attempting to claim deductions for expenses that are not eligible. Failure to report the correct number of employees can be an attempt to evade tax liabilities and other tax-related obligations related to the firms' workforce.

Despite other opportunities to evade, managers may also have stronger incentives or reasons for compliance than employees. For instance, managers may not themselves gain much from non-compliance as they mainly report for others, i.e., their employees and owners, which could partly explain a high compliance rate among managers. Also, firms whose owners care about compliance may hire managers who are likely to be compliant or monitor their compliance more. Managers typically possess a higher level of responsibility and accountability within organizations. Non-compliance with tax obligations may expose the firm to legal risks, including penalties, fines, and potential legal actions. Moreover, managers' tax non-compliance can tarnish the firm's reputation, leading to negative publicity, loss of customer trust, and damage to long-term business relationships. Managers, being responsible for the overall functioning and success of the organization, have a greater stake in safeguarding the firm's legal and reputational standing.

As control variables we add the managers' gender (men as 0 and women as 1), age as a continuous variable and whether they have a foreign background. Managers with a different birth country than Norway are defined as "Foreign". If the information on the birth country is missing, we use citizenship. For foreign managers residence time is calculated based on arrival date to Norway. For Norwegian managers without any registered long-term stays abroad, residence time coincides with age. In the regression models, we do not assume a linear

relationship between residence time and compliance, but divide the sample into three different residence time groups, in line with Bastani et al. (2020), i.e. 0-5 years, 5-10 years and over 10 years. There are no Norwegian managers in the first two residence groups.

Foreign managers' CPI Score gives the score for this person's birth country at the year of arrival to Norway. Norwegian managers were given the CPI Score for Norway in 2018. We want to capture the manager's "last impression" of his or her country of origin at the time of arrival in Norway. To do so, we use the CPI Score of the year of arrival in Norway as a representation of a time-precise image of the tax morale in the country of origin. We assume that any development in this index, following arrival in Norway, will not affect the individual tax morale and that immigrants assimilate the tax morale of Norway with time. Whenever this value is missing for that particular year, we have used the nearest value available. We have also inverted the CPI scale so that a higher number reflects a higher corruption level, as a continuous variable. Testing a CPI dummy where managers with CPI Scores at the same level of Norway or lower are given the value 1, and 0 otherwise, renders no significant CPI-coefficients.

One may argue that the last year in the home country is not representative of the corruption level experience in their country of origin, especially if regime changes or government volatility leads people to leave the country. In that case, a cumulative average or a level at a certain age might be more appropriate. But there are also caveats with such an approach, as an average will not capture recent changes (in either direction), and such changes may influence a person more than their lifetime perception. Nevertheless, CPI scores do not change much over time for the countries in the sample, cf. Appendix Table A1b.

We include armed conflict as a dummy. Conflict is synonymous with the destruction of trust between ethnic, religious or other groups, and institutions typically do not bounce back immediately after a cycle of violence (Collier and Sambanis (2002); Miguel et al. (2011); Bellows and Miguel (2006)). Feldman and Slemrod (2009) find a positive effect of external threats on compliance attitudes, because external threats may affect social identification and patriotism. However, war exposure in their setting is limited to violence *outside* the country of residence and will most likely not reduce trust nor increase trauma in the population, in the way we expect civil war to impose. Lange and Melsom (2022) find a counterintuitive *positive* effect on

employees' compliance from the ratio of employees in the enterprise exposed to armed conflict, and so we include this variable in our OLS models as well.

We use a recovery period of 25 years, i.e. an immigrant exposed to armed conflict from 0 to 25 years prior to registration in Norway, gets the value 1, and 0 otherwise. Thus, we do not measure armed conflicts older than 25 years. This specification follows that of Miguel et al. (2011). The restriction entails that 37.8 percent of the foreign managers in our sample come from countries involved in armed conflicts. Some of these are Western countries, such as the US and the UK, who have intervened in armed conflicts outside their territory. However, the number of managers from these countries is small and unlikely to drive the results. For details on the conflict dummy and which countries it comprises, see Appendix Table A4. We have also tried different versions of the conflict dummy, measuring exposure up to 5, 10, 15 and 20 years prior to migration. None of these yield significant coefficients.

We include the use of an external accountant as a control variable. An external accountant provides an 'arm's length' distance to the manager, and typically has more accounting competence than the latter.

The first stage of the GA and LASSO performs variable selection among the 100 independent variables in our data set. The selected variables from these model runs are included in the summary statistics table. Thus, Table 1 gives summary statistics for variables in all models, broken down by Norwegian versus Foreign Managers. In Appendix Table A2, summary statistics are broken down by compliant versus non-compliant managers.

Table 1: Descriptive statistics

Independent Variables	Norwegian (N=1,580)	Foreign (N=394)	t	P >|t|
Female	0.161	0.224	-2.936	0.003
Age	50.926	44.237	10.951	0.000
Residence Time	51.515	17.612	53.652	0.000
Conflict	0.000	0.378	-25.422	0.000
CPI Score	15.750	49.139	-52.278	0.000
External Accountant	0.786	0.873	-3.889	0.000
Conflict Employees	0.050	0.202	-16.398	0.000
Ltd Company	0.833	0.843	-0.465	0.642
Salary System	0.700	0.683	0.667	0.505
Work Training	0.109	0.112	-0.180	0.857
Job Advertisement	0.216	0.155	2.715	0.007
Time Sheet	0.477	0.531	-0.819	0.413
Terms	10.288	10.251	0.201	0.840
Audit Employees	3.925	4.225	-4.607	0.000
Self-employed	0.163	0.137	11.933	0.000
Dependent Variable (2017)				
Firm Compliance	0.895	0.858	2.084	0.037
Tax deduction	0.950	0.921	2.220	0.026
Payroll accounts	0.965	0.967	-0.177	0.860
Monthly reporting	0.966	0.964	0.132	0.895
General accounts	0.997	0.997	0.002	0.998

Note: Columns 2 and 3 show the mean value of the variable (Column 1) among Norwegian and Foreign managers. Column 4 shows the t-value on the differences between compliant and non-compliant managers. We use the z-test (test of proportion) for binary outcomes and the t-test for continuous outcomes. Column 5 shows the p-value for the test. Female, Conflict, External Accountant, Ltd Company, Salary System, Work Training, Job Advertisement, and Self-employed are dummy variables. Age, Residence Time, CPI Score, Conflict Employees, Time Sheet, Terms and Self-employed are continuous variables. Female takes value 1 if the manager is female. Age value is years. Residence time value is years. Conflict takes the value 1 if the manager was exposed to armed conflict in home country up to 25 years prior to migration. CPI score is the inverted CPI score of the managers' country of origin at the year of arrival to Norway. External accountant takes the value 1 if the firm has outsourced external accountant services. Conflict Employees is the fraction of employees in the firm exposed to armed conflict in the home country up to 25 years prior to migration. Ltd Company takes the value 1 if the firm is registered as a private limited company. Salary system takes the value 1 if the firm has a digital salary system. Work training takes the value 1 if the firm has registered employees in the work training programme subsidised by the Norwegian Labour and Welfare Administration (NAV). Job Advertisement takes the value 1 the firm has advertised for vacancies in Norway. Time Sheet describes the number of employees with incorrect time sheets. Terms describes the number of terms (1-12) the firm report salaries or benefits for any employee to the NTA. Audit employees is the number of audited employees per firm (0-9). Self-employed takes the value 1 if the firm is registered as a sole proprietorship at the NTA.

89.50 per cent of the sample is compliant. In addition, there are only 394 foreign managers who constitute about 20 per cent of the sample. The results on residence time, CPI Score and conflict exposure are thus based on a quite small sample with relatively little variation. Both appropriate sample sizes and sufficient variation in the dependent variables are necessary to provide reliable, reproducible and valid results. Even though a larger sample with more variation is always preferable, our sample is significantly larger than other studies using randomized data (Jenkins and Quintana-Ascencio (2020); Blackford (2017)). It has been difficult to obtain consistent and clear guidelines for minimum N in regression analyses, but the sample size in our study far exceeds the recommendations in a recent study attempting to do so (Jenkins & Quintana-Ascencio, 2020).

22.4 percent of the foreign managers are female, compared to 16.1 percent of the Norwegian managers. 87.3 percent of foreign managers use external accountants, compared to 78.6 percent of Norwegian managers. The fraction of employees exposed to armed conflict is larger in foreign managed firms (20.2 percent) than in Norwegian managed firms (5.0 percent). Company characteristics such as company type, salary system, and work training are quite evenly distributed between the two manager groups. Foreign managers advertise less for vacancies than their Norwegian counterparts.

4. Empirical Strategy

Commonly used dependent variables in studies of tax compliance are differences in self-reported income tax pre- and post-audit (Kleven et al., 2011), tax deficiency to revenue ratio (DeBacker et al., 2015), changes in federal income tax deposits (Boning et al., 2020) or changes in employer-reported income tax (Bjørneby et al., 2021).

Extensive use of third-party reporting and employers' tax withholding are powerful mechanisms to ensure compliance unless the employer and employee collude to evade (Bjørneby et al., 2021). Thus, tax evasion may very well be partly influenced by the employees' decisions, and not solely by the managers. Unlike previous contributions that seek to measure tax evasion through changes in individual tax remittance, we measure managers' tax compliance through their own, direct reporting liabilities. While we cannot infer that variance in these liabilities is due to evasion, this variance is typically not attributable to other parties'

behaviour than the managers. Non-compliance with the reporting requirements of our dependent variable leads to revenue losses through lower remittance of employee tax and payroll tax.

In this paper, we use two strategies for variable selection. First, the "traditional" approach of testing established relationships from the adjacent research literature, using the independent variables from other compliance areas or tax compliance literature. This is the main model. Second, given the large volume of variables available in our sample, we first use a GA to guide the variable selection, and then run a linear regression on a restricted selection. We also run a LASSO model to test whether the LASSO algorithm will select other independent variables. We use both models to validate the results from the OLS model. To obtain inferences relevant to the target population, we will cluster standard errors by stratum (NACE codes) (Solon et al., 2015).

The choice of our ML strategy merits justification. Battaglini et al. (2022) test a random forest model to predict and improve tax auditing efficiency using non-randomized administrative data. While random forest models may also be used for variable selection, we assess that variable selection is not where such models have their advantage. Random forest does not explicitly handle noisy or irrelevant variables. While it can indirectly identify such variables by assigning them lower importance scores, there is no built-in mechanism to explicitly filter out noisy features, as is the case with both GA and LASSO. In some cases, irrelevant variables may still contribute to the variable importance scores and lead to suboptimal selection outcomes. The performance of random forest, including variable importance, can also be sensitive to the choice of hyperparameters such as the number of trees, tree depth, and feature subsampling rate. Suboptimal hyperparameter tuning may affect the accuracy of variable importance rankings and subsequent variable selection decisions.

The data used in this paper provide detailed information on managers' compliance behaviour, firm characteristics, and certain individual characteristics such as age, gender foreign background and residence time in Norway. However, there may be many individual characteristics which may be correlated with compliance but are not observed in these data, and so the estimates will be plagued by omitted variables bias. As an example, the quality of the education is probably positively correlated with a manager’s ability to navigate complicated tax laws (which may both make the manager more likely to be able to comply, but

probably also more likely to be good at evading taxes), and correlated with age, gender, residence time in Norway, and whether the manager comes from a conflict zone. So unobserved variables may confound the associations we observe if they are related to both compliance behaviour and our key independent variables. The Tax Administration has no information on e.g., the managers' field or level of education. Further studies are needed to determine whether there is a causal relationship between residence time, foreign background, and tax compliance. Nevertheless, the supervised ML approach to variable selection allows us to test many variables from the audits which could affect compliance as well, ruling out potential bias from these. The novelty of our contribution is thus that we have information on variables omitted in other papers.

In econometric regression analysis, several hypothesis tests are often performed simultaneously, and this applies to the regressions in this paper as well. The problem then becomes how to decide which hypotheses to reject, or more precisely, whether significant effects that emerge after many different hypothesis tests are real or spurious. Romano and Wolf (2005) propose a step-by-step test procedure which, compared to related test methods, is "more powerful" and will more often reject false hypotheses. We have run Romano-Wolf tests according to the procedure described in Clarke et al. (2019) on all model specifications and found that most corrected p-values as a result of the test are robust. Romano-Wolf implicitly captures common dependency structure in the test statistics, resulting in an increased ability to detect false hypotheses. Hence, the Romano-Wolf gives more robust tests than traditional tests for multiple hypotheses such as Benjamini and Hochberg (1995) or Bonferroni (1936). We include the Romano-Wolf test statistics in the results tables.

4.1 OLS Model Specification

In the OSL model, we explain manager compliance through three independent variables, namely residence time in Norway, CPI Score of the native country upon migration to Norway, and armed conflict exposure in the past 25 years upon migration to Norway. We include the control variables gender, age, foreign-born, fraction of employees exposed to armed conflicts, and the use of an external accountant. We use fixed effect regressions on 4-digit NACE codes to control for differences between industries and, as mentioned we also use fixed effect estimations to control for systematic differences between auditors.

We run one OLS without fixed effects to study the overall effects, irrespective of industry sector variance, and then two OLS with fixed effects, first on the NTA auditor ID, and then on both NTA auditor id and industry sector (Nace code on 2-digit level). To obtain consistent estimates, we cluster standard errors by NACE code (2-digit level). We estimate specifications of the following type:

(1) $Y_i = \beta_1 T_i + \beta_2 CPI_i + \beta_3 War_i + \beta_4 X'_i + \varepsilon_i$

Where the dependent variable is compliance by manager i. T_i is the residence time of that person, CPI_i is his or her CPI Score, and War_i is the exposure to armed conflict. X is a vector of controls, including gender, age, and other variables of interest. As managers are exposed to the Norwegian compliance environment over time, we expect the Time coefficient to be positive. Both CPI and War should influence compliance negatively because they reflect a lack of trust in government, and so we expect the coefficients of these two variables to be negative. There may be problems with potential multicollinearity, which we address by estimating the variance inflation factor (VIF) in the robustness section.

4.2 GA Model Specification

We rely on Broadhurst et al. (1997) in specifying our GA. This is a stochastic optimization technique where a population of n subsets is created, each containing a random variation of variables. Then, the cost function for each subset is sequentially evaluated, creating a new population. We then apply a weighted random selection to the original population, where the probability of a particular subset being selected is a function of its cost function response, i.e., the better the cost function response, the greater the chance of selection. The selection process is repeated until n new subsets are created, their cost functions evaluated and repeated until a stopping criterion is reached.

The GA model consists of two parts. First, a variable selection, where independent variables are selected through a specified number of iterations, and second, it runs a linear regression over these selected variables. We limit the stopping criterion to 8 independent variables. Allowing a higher stopping criterion will, eo ipso increase the explanatory power, but also include variables with very low coefficients, despite them being statistically significant. LASSO deals with this problem through the tuning parameter. We build the model on a training set using .50 of the data and validate the results on the remaining .50. A total of 100 independent

variables (continuous and dummies) from our random audit data set are included in the iterations. We set the number of iterations to 150, as the fitness function does not improve beyond this cut-off, cf. Appendix Figure A1.

The GA model aims to find the best OLS model on the 8-variable subset of the 100 independent variables that we have available. To do so, it randomly generates an initial population of 200 possible models of 8 variables from our 100 available variables and performs the OLS on all of them. The resulting set of 200 8-variable OSL models is then ranked and the most promising ones, with the lowest Mean Squared Error (MSE), crossed to form a new population of 600 8-variable models. The best ones are crossed to form 600 new possible models and the algorithm continues. This algorithm eventually converges to a local maximum of the best OLS model on 8 variables and halts. To explore the different local maxima, the GA algorithm is re-initialized 2,000 times.

The GA algorithm selects the following unique linear model in 754 out of 2,000 runs:

$$(2)\ Y_i = \alpha + \beta_1 X'_i + \varepsilon_i$$

Where the dependent binary variable is compliance by manager i, α is a constant, X is a vector of the following variables: Private limited company; Salary system; Work training; External accountant; Job advertisement; Time Sheet; Terms, and Conflict Employees. ε is the error term. Note that residence time is not selected by the GA, and thus not estimated. The results from the first stage of the model (variable selection) are shown in Appendix Table A5. In the final step, we run this model on the total sample with fixed effects on the NTA auditor ID, and then on the industry sector (Nace code), and cluster standard errors by Nace code (2-digit level).

4.3 LASSO Model Specification

More common than the GA model in economics are Ridge and LASSO regressions (Pereira et al., 2016); (Hansen & Liao, 2019). While Ridge regressions are more suitable in multicollinear data containing a higher number of independent variables than observations, LASSO regressions are suited both for models with high levels of multicollinearity or when we want to automate variable selection, perhaps because no theory is available to guide this selection. In our data set, the observations outnumber the variables by far, so a Ridge model is not suitable. A LASSO model will in our case also fit the purpose, but we assess that the LASSO

entails more preconditions (like "best subset" under a regularized lp-norm (Zhou et al., 2015)) than the GA, so the variable selection is more restricted. The upside of the LASSO compared to the GA is that the latter typically includes some variables with little predictive power in the regression equation, which are likely to be removed in a LASSO model.

Like the GA model, a LASSO also performs variable selection. The results are often easier to interpret than those of a linear regression because the dependent variable will only be explained by a small subset of the predictors, i.e. those with nonzero coefficient estimates (James et al., 2013). The LASSO coefficients minimize the quantity:

$$(3)\ \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{i=0}^{n}\beta_j\, x_{ij}\right)^2 - \lambda \sum_{j=1}^{p}|\beta_j| = RSS + \lambda \sum_{j=1}^{p}|\beta_j|$$

Where $\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{i=0}^{n}\beta_j\, x_{ij}\right)^2$ is the residual sum of squares (RSS), and λ is the tuning parameter determining the punishment of the large coefficients, i.e. if $\lambda = 0$, then the model is equal to a standard OLS. The only hyperparameter we need to determine when running the LASSO algorithm is the λ -value. Following James et al. (2013) we use 10-fold cross-validation optimised for mean square error (MSE) to select the optimal λ-value. However, choosing the folds for the cross-validations introduces randomness in the LASSO algorithm, and hence cross-validation generates different values for optimal λ. To avoid selecting a λ -value at random, we run 2000 10-fold cross-validation and end up with 15 unique values for λ, and hence 15 different models, estimating the effects of from 9 and up to 29 independent variables.

After 2000 runs on 10-fold cross-validation and selecting the model with the closest approximation to the OLS and GA on the number of variables, the resulting model is:

$$(4)\ \mathrm{Y_i} = \alpha + \beta_1 X'_i + \varepsilon_i$$

Where the dependent binary variable is compliance by employer i, α is a constant, X is a vector of the following variables: Private limited company; Self-employed; Salary system; Work training; External accountant; Audited Employees; Time Sheet; Terms, and Conflict Employees. ε is the error term. As for the GA, we run this LASSO model on the total sample with fixed effects on the NTA auditor ID, and then on industry sector (Nace code), and cluster standard errors by Nace code (2-digit level). The results from the LASSO variable selection are shown in Appendix Table A6. xxx

5. Results

5.1 OLS Model results

The results from the OLS model runs are presented in Table 2.

Table 2. Results from the OLS and Fixed Effects models

	(1)	(2)	(3)
Female	0.029	0.033+	0.014
	(0.020)	(0.019)	(0.021)
Age	-0.002***	-0.003***	-0.003***
	(0.001)	(0.000)	(0.000)
Foreign	-0.043+	-0.045+	-0.050*
	(0.024)	(0.024)	(0.023)
<5	-0.121	-0.166	-0.165+
	(0.086)	(0.101)	(0.095)
5-10	-0.025	-0.035	-0.039
	(0.040)	(0.037)	(0.036)
Conflict	-0.005	0.053*	0.056*
	(0.030)	(0.026)	(0.027)
CPI Score	0.001	0.000	-0.000
	(0.001)	(0.000)	(0.000)
External Accountant	0.096***	0.092***	0.092***
	(0.021)	(0.022)	(0.023)
Conflict Employees	-0.108	-0.125	-0.152*
	(0.086)	(0.075)	(0.066)
Constant	0.938***		
	(0.027)		
Observations	1959	1936	1928
R-squared	0.029	0.166	0.236
Romano Wolf Bootstrap p-values	Original	Romano Wolf	
Female	0.069	0.079	
Age	0.001	0.000	
Foreign	0.037	0.050	
Residence Time	0.069	0.069	
Conflict	0.253	0.228	
CPI Score	0.490	0.446	
External Accountant	0.000	0.000	
Conflict Employees	0.042	0.040	

Note: Estimated coefficients from OLS model runs. Standard errors in parentheses. Column (1) is OLS without fixed effects, column (2) is OLS with fixed effects on NTA auditor, and column (3) is OLS with fixed effects on NTA auditor and NACE code (2-digit). Residence time >10 years is the reference category and therefore omitted in the table. Standard errors are clustered by NACE code. Romano Wolf test statistics are given for model (3) by the original and Romano Wolf p-values of each independent variable. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

Foreign managers are 5 percent less likely to reach full compliance (3). Bastani et al. (2020) reveal that immigrants are more likely to miss the declaration deadline and be fined for non-compliance, regardless of their residence time in Sweden, but no such effects are replicated in this context. Part of their explanation is the language barrier among immigrants, which may be lower among non-native managers in our sample. On the contrary, we find only a significant and negative effect among the most recently arrived (<5), suggesting a learning effect after 5 years of stay. We find lower coefficients for those managers with residence time < 5 years than those with 5-10 years, compared to those with >10 years of stay in Norway. This tendency brings again some associations to Bastani et al. (2020) who find that the probability of taking up the commuting deduction in the Swedish tax system is lower among immigrants with residence time < 5 years than those with 5-10 years of stay in Sweden, compared to Swedish natives. As the significance level of our result is weak, one should be careful to suggest common influence from confounding variable(s), even though the patterns appear similar to that of Bastani et al. (2020).

We find a negative association between employee conflict exposure and managers' compliance (3). While Lange and Melsom (2022) find a *positive* effect on employees' compliance from other peer employee conflict exposure, the effect on manager compliance is negative. A 10 percent increase in the fraction of employees exposed to armed conflict is associated with a 15.2 percent decrease in the probability of management compliance (3). As a robustness test, we have run all model specifications with each of the four components of the dependent variable. From Appendix Table A7, we see that only the component "Tax deduction" is driving this result, i.e. a 10 percent increase in the fraction of employees exposed to armed conflict is associated with a 14.9 percent decrease in the probability of correct tax deduction filing by the manager. One explanation could be selection. Firms with many marginalized employees, such as war refugees, may have other characteristics that result in poorer compliance, i.e. higher staff turnover or less experienced managers.

We find no stable associations between the managers' conflict exposure and compliance, and no effects from CPI Score in this sample. Fisman and Miguel (2007) find that home country

corruption norms are an important predictor of the propensity to behave corruptly among diplomats, but the cultural mechanism explaining non-compliance, i.e., unpaid parking violations, in their sample, seems not to be at play in our case. One explanation may be that codes of conduct in the workplace environment 'eradicate' or 'neutralize' cultural background characteristics such as corruption because managers are exposed to tax reporting standards as an integrated part of their everyday business. Diplomats' parking routines, or lack thereof, on the other hand, are not an integrated part of their occupations as diplomats, but rather their character. Indeed, Fisman and Miguel (2007) find that the time and space of many violations is strong evidence that these are not even work-related, and that third-party reporting closes the incentives to misreport.

We find a small negative, but statistically significant effect of age on management compliance across all 3 model specifications. The age effect is stable, i.e., the probability of management compliance *decreases* with a ratio of .003 (3) as the manager gets one year older. Nevertheless, the sign of the age coefficient may not be surprising as age correlates with seniority in any position, and thus the more senior, the more knowledge about the loopholes in the tax system. Furthermore, perhaps managers over time get a more realistic view of the (low) probability of audit selection by the tax authorities. This effect may also be explained by a higher understanding of tax legislation among managers in our sample, than in the population in general. The age-learning effect may be equivalent to the age gradient identified by Bastani et al. (2020). They find an age gradient in the take-up of commuter deductions for natives and immigrants with long residence time in Sweden, who presumably have adapted more to the tax system compared to newly arrived and younger immigrants. Furthermore, this learning effect resembles a learning effect in Fisman and Miguel (2007), where diplomats become bolder in their violations once they "successfully "got away with it" a few times (or heard stories about others doing so)." (Fisman & Miguel, 2007, p. 1042)

However, the negative age effect partly contradicts more recent findings. Hofmann et al. (2017) conducted a comprehensive meta-study on tax compliance across sociodemographic categories including age, and found a small positive, but significant relation between the age of taxpayers and their tax compliance. This confirms findings in Nordblom and Žamac (2012) and Kirchler (2007), as well as older studies, such as Tittle (1980), Witte and Woodbury (1985), Dubin and Wilde (1988), Feinstein (1991), and Hanno and Violette (1996). Ashby et al. (2009), Braithwaite and Ahmed (2005), and Muehlbacher et al. (2011) find no age effect, however.

Furthermore, we find a significant positive effect of holding an external accountant across all model specifications. In column (3) of Table 2 we observe that hiring an external accountant is associated with an increase in compliance with a ratio of .092 when controlled for the auditor and industry sector. Most of the effect stems from the mandatory monthly reporting, cf. Appendix Table A7. The effect is expected, as one would typically infer that accountants possess more knowledge of tax filing and liability than business managers in particular and people in general. The positive effect may both be causal and due to selection. Using an external accountant means hiring someone who is authorized to keep accounts and is obliged to ensure compliance with rules and regulations. In addition, using an external accountant is not mandatory. Thus it seems likely that firms more concerned with compliance will do so to a greater extent than other firms. Saad (2014) finds some degree of trust in accountants' tax knowledge as reasons for outsourcing tax filing. Further to this, managers with different characteristics and compliance attitudes may be selected to different sectors. However, adding fixed effects on Nace code (2-digit) on the third model specification, does not alter the significance level, and so this prospective selection bias is unlikely to drive the results.

For foreign managers, the residence time is calculated based on the arrival date to Norway. For Norwegian managers without any registered long-term stays abroad, residence time coincides with age. As about 80 per cent of the managers are Norwegian, age and residence time are highly correlated in the full sample. In addition, residence time has a slightly different meaning for foreign and Norwegian managers. To address these issues, we have tested interaction terms between residence time and foreign background to see if the association between residence time is different for Norwegian and foreign managers but found no significant interaction effects. We have also tried to run the regressions on residence time separately on Norwegian and foreign workers. Limiting the sample to Norwegian managers, the variables on CPI, conflict and residence time are omitted, leaving only gender, age and external accountant as independent variables. The coefficient for age is still significant for Norwegian managers. For foreign managers, neither age nor residence time yields significant coefficients.

Conflict and corruption variables are also correlated. The average CPI Score is not surprisingly substantially higher in countries with records of armed conflict. To check this association more closely we have regressed Compliance over these two variables separately and then combined.

In these models the coefficients remain stable in all models, cf. Appendix Table A10. This indicates that CPI Scores measure something different than armed conflict and we thus argue that it is possible to separate their association with compliance from one another.

There is also considerable variation in CPI Scores in both groups. For managers from countries exposed to armed conflict, the CPI Score ranges from 12 to 85. For managers from non-conflict countries, it ranges from .6 to 80. Except for the vast majority of Norwegian managers with the CPI Score for Norway in 2018, the CPI Scores are quite evenly distributed in both groups.

5.2 GA Model Results

The results from the GA model specification are displayed in Table 3.

Table 3. GA Model results

	(1)	(2)	(3)
Ltd Company	0.132***	0.145***	0.103***
	(0.030)	(0.029)	(0.022)
Salary system	0.085**	0.108***	0.103***
	(0.025)	(0.029)	(0.029)
Work training	0.037+	0.028+	0.024
	(0.020)	(0.017)	(0.017)
External Accountant	0.129***	0.115***	0.111***
	(0.018)	(0.020)	(0.021)
Job advertisement	0.024*	0.024**	0.019+
	(0.010)	(0.009)	(0.010)
Time Sheet	0.003	0.000	0.003
	(0.005)	(0.006)	(0.007)
Terms	0.022***	0.023***	0.021***
	(0.003)	(0.003)	(0.003)
Conflict Employees	-0.126**	-0.094*	-0.119*
	(0.047)	(0.041)	(0.048)
Constant	0.390***		
	(0.078)		
Observations	1897	1872	1864
R-squared	0.154	0.283	0.313
Romano Wolf Bootstrap p-values	Original	Romano Wolf	
Ltd Company	0.000	0.000	
Salary system	0.000	0.000	
Work training	0.000	0.000	
External accountant	0.000	0.000	
Job advertisement	0.000	0.000	
Time Sheet	0.479	0.475	
Terms	0.000	0.000	
Conflict Employees	0.042	0.050	

Note: Estimated coefficients from the OLS model runs, after GA variable selection. Standard errors in parentheses. Column (1) is OLS without fixed effects, column (2) is OLS with fixed effects on NTA auditor, and column (3) is OLS with fixed effects on NTA auditor and NACE code (2-digit). Standard errors are clustered by NACE code. Romano Wolf test statistics are given for model (3) by the original and Romano Wolf p-values of each independent variable. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

We observe that holding a Private limited company increases the probability of compliance by a ratio of .103 (3), and using an external accountant increases the probability of compliance by a ratio of .111 (3). There is a positive effect of .021 (3) on firm compliance of Terms. A firm reporting salaries or benefits for employees has sustained activity and therefore reporting liabilities, but, perhaps equally important, more contact points with NTA throughout the year. The negative effect from the fraction of firm employees exposed to armed conflict is reproduced in this model, although slightly weaker.

The returns from the GA model yield intuitive results. A private limited company is more transparent, has more reporting liabilities and is under easier surveillance and scrutiny by the tax authorities than self-employed or registered foreign companies. Thus, one would expect a higher probability of compliance for private limited companies. The use of an external accountant also increases the probability of compliance, and this coefficient replicates the effects from the OLS models. Most of the effects of holding an external accountant is driven by monthly reporting, cf. Appendix Table A7-A9.

5.3 LASSO Model Results

The results from the LASSO model specification are displayed in Table 4.

Table 4. LASSO Model results

	(1)	(2)	(3)
Ltd Company	0.108	0.131	0.145
	(0.113)	(0.104)	(0.103)
Self-employed	-0.026	-0.014	0.050
	(0.118)	(0.108)	(0.107)
Salary system	0.085**	0.107***	0.102***
	(0.025)	(0.029)	(0.029)
Work training	0.039+	0.030+	0.024
	(0.020)	(0.016)	(0.017)
External Accountant	0.128***	0.115***	0.113***
	(0.019)	(0.021)	(0.023)
Audit Employees	0.003	0.005	0.010*
	(0.004)	(0.004)	(0.004)
Time Sheet	0.003	-0.001	0.001
	(0.005)	(0.006)	(0.006)
Terms	0.022***	0.022***	0.020***
	(0.004)	(0.003)	(0.003)
Conflict Employees	-0.127**	-0.096*	-0.124*
	(0.046)	(0.040)	(0.047)
Constant	0.408***		
	(0.112)		
Observations	1897	1872	1864
R-squared	0.153	0.283	0.314
Romano Wolf Bootstrap p-values	Original	Romano Wolf	
Ltd Company	0.000	0.000	
Self-employed	0.000	0.000	
Salary system	0.000	0.000	
Work training	0.000	0.000	
External Accountant	0.000	0.000	
Audit Employees	0.000	0.000	
Time Sheet	0.479	0.495	
Terms	0.000	0.000	
Conflict Employees	0.042	0.040	

Note: Estimated coefficients from OLS model runs, after LASSO variable selection. Standard errors in parentheses. Column (1) is OLS without fixed effects, column (2) is OLS with fixed effects on NTA auditor, and column (3) is OLS with fixed effects on NTA auditor and NACE code (2-digit). Standard errors are clustered by NACE code. Romano Wolf test statistics are given for model (3) by the original and Romano Wolf p-values of each independent variable. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

The LASSO model results by and large reproduce the effects from the GA model, except for the significant effect of holding a private limited company. Neither company type Ltd Company nor Self-employed are significant predictors of management compliance in this model. This variance between the algorithms is not surprising, however. The inconsistency between the two algorithms is expected since the two algorithms optimize different functions defined over different spaces. The LASSO algorithm optimizes a carefully modified linear regression problem, while the GA algorithm optimizes over the space of all possible linear regressions.

These different goals mean that there is neither reason to expect the algorithms to agree about everything, nor an indication of an arbitrary result. The 'surprise' is rather that the GA and LASSO agree on 7 out of 8 variables and the low variance in size of the coefficients between the two.

Appendix Table A11 gives the results of the OLS, GA, and LASSO model runs with fixed effects on NTA auditor and Nace code (2-digit) in one table.

5.4 GA and LASSO Model Performance Comparison

There is several predictive performance indicators in the literature such as the Akaike Information Criterion (Sakamoto et al., 1986), the Bayesian Information Criterion (Watanabe, 2013), or Adjusted R-squared (Mullainathan & Spiess, 2017). The latter demonstrates the need for a hold-out sample to assess performance. Certain ML algorithms' tendency to overfit is also prevalent in our GA model. Thus, one may expect performance to be overstated in the training sample. A second lesson to learn from Mullainathan and Spiess (2017) is that ML algorithms can perform significantly better than OLS, even when sample sizes and number of covariates are limited. In our setting, it makes less sense to compare the performance of predictive models with the OLS, as we are mainly using the ML models to guide variable selection. However, as we want to check how well the ML models' predictions match the observed data, we find Mean Squared Error (MSE) to be appropriate (James et al., 2013). The performance of the models is displayed in Table 5, where we also include a column for Adjusted R-squared for illustration purposes.

Table 5. Model Performance Comparison

Model	MSE	Adjusted R^2
GA OLS	0.0748	0.1501
GA Auditor	0.0705	0.2833
GA Auditor NACE	0.0699	0.2146
LASSSO OLS	0.0749	0.1490
LASSO Auditor	0.0706	0.2041
LASSO Auditor NACE	0.0698	0.2157

Note: Mean Squared Error and R-Squared for the GA and LASSO models.

Although the ML models have their obvious limitations concerning causal inference (Pearl, 2018), when lack of previous empirical findings or theory cannot guide any explanations of the relationships, there are still lessons to be drawn from a comparison between who ML models performing variable selection in the first stage. We see that the performance is very similar, and so the choice of models should be guided by other criteria, like e.g., the number of parameters one would have to 'arbitrarily' set.

5.5 Further Robustness Tests

To get a clearer picture of the origins of significant effects, we have run all model specifications on each component of the dependent variable, namely tax deduction, payroll accounts, monthly reporting and general accounts. An unambiguous result of this test is that most of the effects of holding an external accountant are driven by monthly reporting. All results are given in Appendix Table A7-A9.

As NTA auditors have suggested independent variables, some multicollinearity between variables likely exist. We estimate variance inflation factor (VIF) of the individual variables in all regressions and find that none of the variables in the standard OLS regressions have a higher VIF than 4.22, and none in the GA- and LASSO-specified regressions have a higher VIF than 1.21, and hence we can disregard multicollinearity.

6. Conclusion

We find a small, negative association between managers' age and compliance, but positive associations between the use of an external accountant, and compliance with reporting requirements. Whereas exposure to armed conflict among the employees in the firm also reduces compliance, our findings from the ML models suggest positive associations between compliance and company characteristics such as holding a private limited company, and internal firm characteristics like an established salary system and the frequency of terms with salary or benefit payments to employees. We find no associations between managers' own conflict exposure and compliance.

The OLS, the GA and the LASSO models all show higher compliance among managers who use an external accountant. However, the use of an external accountant is most likely an endogenous variable. Unlike true independent variables such as gender and age, the use of an external accountant is a choice the managers make. Whether the manager chooses to use an external accountant or not may also be seen as a part of their compliance behaviour. It is not surprising that managers with external accounts are more compliant. However, it is difficult to determine whether this is a result of the managers' inherent inclination to comply, or the services provided by the external accountant. More research is needed to establish causal relationships.

We find no evidence that the home country corruption level has any effect on compliance, nor a clear effect of residence time, except that managers with <5 years length of stay in Norway are less compliant than those with residence time >10 years, suggesting a "learning effect" after 5 years of stay.

The negative age effect in the OLS models partly contradicts recent findings. Thus, the age effect is likely more context and application specific. The positive sign of the coefficient on external accountant is expected, as we infer that accountants possess more advanced knowledge on tax filing and liability than business managers in particular and people in general.

Allowing for non-parametric inference, using the pool of 100 prospective variables suggested by NTA tax auditors, a second contribution of this paper is to test two ML algorithms, namely GA and LASSO. We find that both the GA and LASSO model specification selects other

independent variables than the standard OLS, except for the use of an external accountant and the fraction of employees exposed to armed conflict.

The policy implications for the Tax Administration are twofold. First, the models in this paper have produced significant explanations on some factors driving manager compliance, namely holding an external accountant, having a salary system, and managing a private limited company. These factors contribute positively to manager compliance. Age and previous armed conflict exposure of the employees in the firm, on the other hand, contribute negatively to manager compliance. Audit selection should thus take these characteristics into account when limited audit resources are allocated.

Second, when the Tax Administration possesses representative data, supervised machine learning models such as the GA and LASSO may provide useful tools in both understanding the drivers behind non-compliance and guiding audit selection.

The results of our analysis hint at the use of individual-level variables to target audits. On one hand, this might lead to an improvement in audit measures (whatever is used). On the other hand, it might lead to a discriminated treatment against a particular socio-demographic group by a state institution. Therefore, we place more confidence in the ML results, which tend to uncover firm-level variables.

The decision of whom to select for an audit is based on a comprehensive assessment in which information about entirely different factors is certainly more important. Nevertheless, knowledge about the characteristics that distinguish compliant from non-compliant businesses can be useful, even when considering these types of characteristics. It can provide relevant information when selecting audit targets, but more importantly, it can be valuable knowledge when designing other measures and initiatives to enhance compliance.

7. Appendix

Table A1a: The population of businesses

Criterion
1. Total revenue in 2017 > NOK 100,000
2. If not 1, then total revenue in 2016 ≥ NOK 200,000.
3. Mean active work relations per month is ≤ 20
4. Mean active employees per month is ≥ 5
5. If not 4, then registered > NOK 100,000 on subcontracts and/or foreign services in 2016
6. Private Limited Companies (AS) self-employed (ENK) and Norwegian foreign-registered enterprises (NUF)
7. No termination date registered
8. Sectors in public service, defence and public social security are excluded

Note: Criterion 3 was set to avoid including large (compliant) corporates

Table A1b: CPI Score sample statistics

Country	Min	Max	Mean	Median	STD	N
Afghanistan	8	27	16	16	6.662	17
Algeria	23	36	29	29	4.851	20
Armenia	25	43	35	35	4.054	16
Australia	77	89	82	82	3.530	25
Austria	69	97	81	76	9.730	25
Belarus	20	48	32	32	8.505	21
Bosnia and Herzegovina	25	42	32	32	6.327	20
Bulgaria	29	43	37	41	4.754	17
Canada	73	92	82	81	5.171	25
Chile	50	79	65	67	7.963	19
China	22	53	38	39	6.085	22
Colombia	22	40	36	37	3.761	23
Croatia	27	52	46	49	6.040	22
Czech Republic	18	59	41	48	12.775	22
Denmark	87	100	93	93	3.510	25
Egypt	26	37	31	32	3.033	19
El Salvador	22	39	32	35	5.964	18
Eritrea	18	30	23	23	3.491	20
Estonia	57	80	71	73	6.041	22
Ethiopia	17	37	27	28	6.446	20
Finland	85	100	92	91	4.328	25
France	51	72	66	69	5.987	24
Germany	78	90	84	82	4.231	25
Ghana	27	48	36	34	7.034	22
Greece	26	54	38	38	8.019	25
Hong Kong	70	87	78	77	3.933	25
Hungary	35	55	44	44	6.468	25
Iceland	76	93	82	82	5.047	22
India	26	47	34	35	4.924	24
Iran	25	36	28	28	2.678	17
Iraq	16	22	19	18	1.886	17
Ireland	69	86	75	75	4.260	25
Italy	30	57	45	45	5.951	25
Kazakhstan	20	34	25	25	3.679	21
Korea, South	34	59	46	47	7.591	25
Kosovo	32	39	34	34	1.987	20
Latvia	27	71	57	60	11.411	22
Lebanon	21	30	26	27	2.971	20
Libya	14	28	20	20	3.774	20
Lithuania	38	60	51	53	6.694	21
Montenegro	41	53	47	48	3.838	20

Morocco	33	43	37	37	2.641	22
Netherlands	82	92	86	87	2.960	25
New Zealand	87	96	93	94	2.633	25
Nigeria	7	28	20	24	6.168	23
Norway	84	94	89	89	3.331	25
Pakistan	10	33	26	27	5.087	19
Philippines	23	38	32	34	4.165	18
Poland	42	65	57	60	6.186	16
Romania	30	63	47	47	8.986	20
Russia	16	29	24	24	3.929	21
Saudi Arabia	39	55	48	49	3.889	20
Serbia	24	42	33	37	6.660	18
Slovakia	35	51	45	46	4.666	22
Spain	43	76	61	61	7.012	25
Sri Lanka	25	40	33	37	6.054	15
Sweden	84	95	91	92	3.171	25
Switzerland	79	91	87	86	2.210	25
Syria	13	34	22	24	7.215	17
Thailand	19	38	28	30	6.801	25
Turkey	29	50	37	36	6.517	22
Ukraine	22	32	27	26	2.449	18
United Kingdom	74	88	83	85	4.403	25
United States of America	63	78	72	73	3.607	25
Vietnam	25	37	32	33	3.247	22

Note: Columns 2-4 show the minimum, maximum, mean and median CPI Scores of the countries represented in the sample. Column 5 shows the standard deviation, and column 6 shows the number of CPI Scores registered for each country in the years 1995-2019.

Table A2: Descriptive statistics by compliance

Independent Variables	Non-Compliant (N=222)	Compliant (N=1,752)	t	P >|t|
Female	0.129	0.179	-1.818	0.069
Age	52.055	49.330	3.452	0.001
Foreign	0.252	0.193	2.084	0.037
Residence Time	46.000	44.826	0.939	0.348
Conflict	0.095	0.073	1.144	0.252
CPI Score	22.973	22.114	0.690	0.490
External Accountant	0.667	0.821	-5.444	0.000
Conflict Employees	0.102	0.077	2.035	0.042
Ltd Company	0.550	0.871	-12.152	0.000
Salary System	0.437	0.729	-8.931	0.000
Work Training	0.036	0.119	-3.693	0.000
Job Advertisement	0.095	0.218	-4.299	0.000
Time Sheet	0.430	0.494	-0.708	0.479
Terms	6.977	10.699	-17.317	0.000
Audit Employees	3.516	4.037	-4.607	0.000
Self-employed	0.432	0.123	11.933	0.000

Note: Columns 2 and 3 show the mean value of the variable (Column 1) among compliant and non-compliant managers. Column 4 shows the t-value on the differences between compliant and non-compliant managers. We use the z-test (test of proportion) for binary outcomes and the t-test for continuous outcomes. Column 5 shows the p-value for the test. Female, Foreign, Conflict, External Accountant, Ltd Company, Salary System, Work Training, Job Advertisement, and Self-employed are dummy variables. Age, Residence Time, CPI Score, Conflict Employees, Time Sheet, Terms and Self-employed are continuous variables. Female takes value 1 if the manager is female. Age value is years. Foreign takes value 1 if the manager is foreign. Residence time value is years. Conflict takes the value 1 if the manager was exposed to armed conflict in the home country up to 25 years prior to migration. CPI score is the inverted CPI score of the managers' country of origin at the year of arrival to Norway. External accountant takes the value 1 if the firm has outsourced external accountant services. Conflict Employees is the fraction of employees in the firm exposed to armed conflict in the home country up to 25 years prior to migration. Ltd Company takes the value 1 if the firm is registered as a private limited company. Salary system takes the value 1 if the firm has a digital salary system. Work training takes the value 1 if the firm has registered employees in the work training programme subsidised by the Norwegian Labour and Welfare Administration (NAV). Job Advertisement takes the value 1 the firm has advertised for vacancies in Norway. Time Sheet describes the number of employees with incorrect time sheets. Terms describes number of terms (1-12) the firm reports salaries or benefits for any employee to the NTA. Audit employees is the number of audited employees per firm (0-9). Self-employed takes the value 1 if the firm is registered as a sole proprietorship at the NTA.

Table A3: Represented Sectors by NACE codes

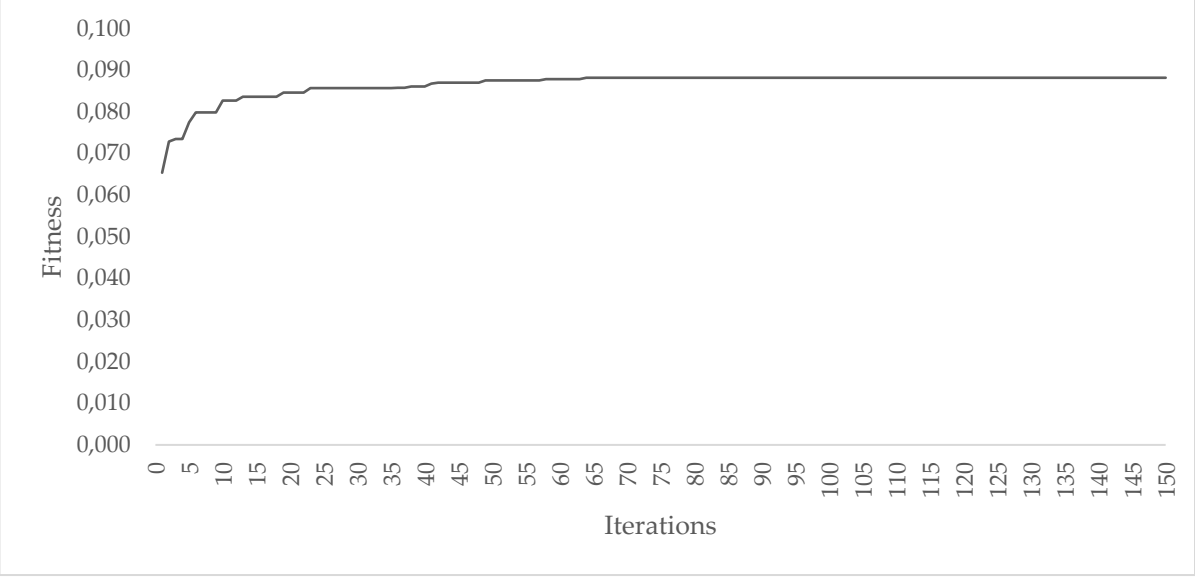
NACE	Description
H49.3.9	Other passenger land transport n.e.c.
B	Mining and quarrying
G47	Retail trade, except of motor vehicles and motorcycles
D35.3	Steam and air conditioning supply
A3	Fishing and aquaculture
N	Administrative and support service activities, excl. N81.2 - Cleaning activities
S96.0.2	Hairdressing and other beauty treatment
S96.0.4	Physical well-being activities
Q	Human health and social work activities
C	Manufacturing
J	Information and communication
A1	Crop and animal production, hunting and related service activities
R	Arts, entertainment, and recreation
H49	General land transport (mainly freight)
F43.341	Paints, coatings, (other finishing without carpentry)
F41	Construction of buildings
F42	Civil engineering
I55	Accommodation
M69.2	Accounting, bookkeeping and auditing activities; tax consultancy
N81.2	Cleaning activities
I56	Food and beverage service activities
F43.3.2	Joinery installation
H	Transport and storage excluding land transport
E	Water supply; sewerage; waste management and remediation activities
G45.2	Maintenance and repair of motor vehicles
	Missing Nace code

Table A4: Conflict Frequency

Country	(1)	(2)	(3)
Turkey	21	14.09	14.09
Pakistan	17	11.41	25.50
Iraq	12	8.05	33.56
China	10	6.71	40.27
Serbia	10	6.71	46.98
India	9	6.04	53.02
Iran	8	5.37	58.39
Sri Lanka	8	5.37	63.76
Vietnam	8	5.37	69.13
Thailand	7	4.70	73.83
USA	4	4.70	78.52
Ethiopia	3	2.68	81.21
Afghanistan	3	2.01	83.22
Lebanon	3	2.01	85.23
Macedonia	3	2.01	87.25
UK	3	2.01	89.26
Eritrea	2	1.34	90.60
Morocco	2	1.34	91.95
Russia	2	1.34	93.29
Algeria	1	0.67	93.96
Australia	1	0.67	94.63
Columbia	1	0.67	95.30
Egypt	1	0.67	95.97
El Salvador	1	0.67	96.64
Philippines	1	0.67	97.32
Libya	1	0.67	97.99
Niger	1	0.67	98.66
Syria	1	0.67	99.33
Ukraine	1	0.67	100.00

Note: Column (1) is frequency, column (2) is percentage of sample, and column 3 is cumulative average.

Figure A1: GA Model fitness function



Note: Fitness is adjusted for R-squared on all variables.

Table A5: GA Variable selection results

Variable	Training		Validation	
Ltd Company	1.159e-01	***	1.541e-01	***
	2.480e-02		2.680e-02	
Salary system	7.969e-02	***	1.093e-01	***
	1.958e-02		2.039e-02	
Work training	5.897e-10	**	-4.568e-11	
	1.964e-10		1.655e-10	
External Accountant	1.059e-01	***	1.608e-01	***
	2.295e-02		2.282e-02	
Job advertisement	6.206e-02	**	-6.847e-03	
	2.289e-02		2.250e-02	
Time Sheet	1.710e-10	**	1.339e-10	**
	5.556e-11		5.012e-11	
Terms	2.301e-02	***	2.447e-02	***
	3.226e-03		3.282e-03	
Conflict Employees	-1.115e-01	**	-8.199e-02	+
	3.992e-02		4.341e-02	
Constant	4.193e-01	***	3.078e-01	***
	4.358e-02		4.094e-02	

Note: Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

Table A6: LASSO Variable selection results

Variable	Training		Validation	
Ltd Company	1.103e-01		1.360e-01	
	1.131e-01		9.652e-02	
Self-employed	-1.215e-02		-1.750e-02	
	1.151e-01		9.861e-02	
Salary system	8.072e-02	***	1.088e-01	***
	1.966e-02		2.040e-02	
Work training	5.962e-10	**	-4.186e-11	
	1.973e-10		1.656e-10	
External Accountant	9.407e-02	***	1.619e-01	***
	2.297e-02		2.291e-02	
Audit Employees	-3.997e-03		4.075e-03	
	5.409e-03		5.717e-03	
Time Sheet	3.997e-03		-4.075e-03	
	5.409e-03		5.717e-03	
Terms	2.454e-02	***	2.394e-02	***
	3.286e-03		3.313e-03	
Conflict Employees	-1.169e-01	**	-8.007e-02	+
	4.006e-02		4.351e-02	
Constant	4.448e-01	***	3.147e-01	**
	1.203e-01		1.023e-01	

Note: Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

Table A7: Results from the OLS Fixed Effects models on 4 dependent variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female	0.013	0.015	-0.005	0.022*	0.022+	0.020	-0.001	0.001	0.003	0.003*	0.003+	0.003
	(0.011)	(0.010)	(0.012)	(0.009)	(0.012)	(0.014)	(0.010)	(0.010)	(0.011)	(0.001)	(0.002)	(0.002)
Age	-0.001*	-0.001**	-0.001*	-0.001***	-0.001**	-0.001**	-0.001+	-0.001*	-0.001*	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Foreign=1	-0.031	-0.032	-0.034	0.018*	0.022**	0.026**	-0.015	-0.018	-0.025	-0.002	-0.003	-0.002
	(0.026)	(0.031)	(0.028)	(0.009)	(0.008)	(0.009)	(0.020)	(0.016)	(0.015)	(0.004)	(0.005)	(0.005)
<5	-0.014	-0.034	-0.043	-0.112+	-0.127*	-0.126*	-0.003	-0.021	-0.017	-0.036	-0.028	-0.030
	(0.066)	(0.073)	(0.068)	(0.066)	(0.063)	(0.060)	(0.034)	(0.038)	(0.040)	(0.033)	(0.037)	(0.039)
5-10	-0.027	-0.032	-0.036	-0.005	-0.015	-0.012	-0.019	-0.025	-0.025	0.001	-0.000	-0.001
	(0.030)	(0.031)	(0.029)	(0.015)	(0.021)	(0.023)	(0.029)	(0.025)	(0.025)	(0.001)	(0.002)	(0.002)
Conflict	0.085**	0.103**	0.099**	-0.038+	-0.004	-0.003	-0.030	-0.022	-0.011	-0.004	-0.003+	-0.003+
	(0.027)	(0.034)	(0.028)	(0.022)	(0.021)	(0.022)	(0.022)	(0.023)	(0.024)	(0.003)	(0.002)	(0.001)
CPI Score	-0.000	-0.001	-0.001+	0.000	-0.000	-0.000	0.001	0.001*	0.001+	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
External Accountant	0.018	0.021+	0.025+	0.004	0.011	0.010	0.083***	0.072***	0.070***	0.013*	0.012*	0.013*
	(0.012)	(0.012)	(0.013)	(0.009)	(0.010)	(0.010)	(0.016)	(0.016)	(0.016)	(0.006)	(0.006)	(0.006)
Conflict Employees	-0.122+	-0.128*	-0.149**	-0.002	-0.005	-0.005	-0.009	-0.014	-0.024	0.003	0.004	0.007+
	(0.071)	(0.062)	(0.055)	(0.026)	(0.024)	(0.022)	(0.028)	(0.031)	(0.033)	(0.002)	(0.003)	(0.004)
Constant	1.007***			1.030***			0.920***			0.990***		
	(0.018)			(0.016)			(0.021)			(0.007)		
Observations	1959	1936	1928	1959	1936	1928	1959	1936	1928	1959	1936	1928
R-squared	0.013	0.139	0.223	0.015	0.211	0.240	0.035	0.171	0.203	0.019	0.085	0.095

Note: Estimated coefficients from OLS model runs on Tax deduction (1-3); Payroll accounts (4-6); Monthly reporting (7-9) and General accounts (10-12). Standard errors in parentheses. Column (1, 4, 7 and 10) is OLS without fixed effects, column (2, 5 and 11) is OLS with fixed effects on NTA auditor, and column (3, 6 and 12) is OLS with fixed effects on NTA auditor and Nace code (2-digit). Residence time >10 years is the reference category and therefore omitted in the table. Standard errors are clustered by Nace code. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

Table A8: Results from the GA models on 4 dependent variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ltd Company	0.160***	0.157***	0.123***	-0.002	0.003	0.004	0.009	0.011	0.004	-0.001	-0.001	-0.001
	(0.029)	(0.026)	(0.023)	(0.010)	(0.011)	(0.014)	(0.011)	(0.010)	(0.012)	(0.004)	(0.003)	(0.005)
Salary system	0.027	0.030	0.029	0.064***	0.095***	0.096***	0.023+	0.029+	0.026+	0.002	0.002	0.001
	(0.019)	(0.020)	(0.020)	(0.010)	(0.015)	(0.015)	(0.011)	(0.016)	(0.016)	(0.003)	(0.005)	(0.006)
Work training	0.019	0.010	0.008	0.010+	-0.003	-0.006	0.012	0.021*	0.019*	-0.004	-0.004	-0.004
	(0.014)	(0.012)	(0.013)	(0.006)	(0.005)	(0.006)	(0.010)	(0.009)	(0.009)	(0.004)	(0.004)	(0.004)
External Accountant	0.041**	0.033*	0.034*	0.019**	0.025**	0.024*	0.091***	0.080***	0.078***	0.014*	0.014*	0.015*
	(0.013)	(0.013)	(0.014)	(0.007)	(0.009)	(0.009)	(0.018)	(0.018)	(0.018)	(0.006)	(0.006)	(0.006)
Job advertisement	0.012*	0.014*	0.003	0.009+	0.009	0.007	0.007	0.003	0.012	0.004+	0.006*	0.006*
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)	(0.007)	(0.008)	(0.002)	(0.002)	(0.003)
Time Sheet	0.003	0.002	0.002	0.003+	0.002	0.001	0.001	-0.002	0.001	-0.000	0.001	0.000
	(0.005)	(0.006)	(0.006)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.000)	(0.001)	(0.001)
Terms	0.016***	0.016***	0.015***	0.010***	0.009***	0.008***	0.002	0.003*	0.003	0.001	0.001	0.001
	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Conflict Employees	-0.066*	-0.056*	-0.074**	-0.051*	-0.026	-0.025	-0.018	-0.011	-0.017	0.004*	0.005	0.010+
	(0.026)	(0.025)	(0.025)	(0.023)	(0.022)	(0.023)	(0.022)	(0.023)	(0.029)	(0.002)	(0.005)	(0.006)
Constant	0.594***			0.812***			0.843***			0.974***		
	(0.076)			(0.030)			(0.025)			(0.014)		
Observations	1897	1872	1864	1897	1872	1864	1897	1872	1864	1897	1872	1864
R-squared	0.161	0.261	0.286	0.103	0.335	0.350	0.041	0.187	0.220	0.016	0.088	0.099

Note: Estimated coefficients from LASSO model runs on Tax deduction (1-3); Payroll accounts (4-6); Monthly reporting (7-9) and general accounts (10-12). Standard errors in parentheses. Column (1, 4, 7 and 10) is OLS without fixed effects, column (2, 5 and 11) is OLS with fixed effects on NTA auditor, and column (3, 6 and 12) is OLS with fixed effects on NTA auditor and Nace code (2-digit). Residence time >10 years is the reference category and therefore omitted in the table. Standard errors are clustered by Nace code. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

Table A9: Results from the LASSO models on 4 dependent variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ltd Company	0.099	0.115	0.124	-0.025*	-0.009	-0.000	0.022	0.024	0.026	-0.007*	-0.004	-0.003
	(0.093)	(0.092)	(0.093)	(0.012)	(0.012)	(0.012)	(0.076)	(0.058)	(0.061)	(0.003)	(0.003)	(0.003)
Self-employed	-0.063	-0.043	0.005	-0.023	-0.011	-0.000	0.011	0.011	0.023	-0.006	-0.003	-0.002
	(0.095)	(0.095)	(0.097)	(0.014)	(0.019)	(0.020)	(0.079)	(0.063)	(0.066)	(0.004)	(0.003)	(0.005)
Salary system	0.027	0.029	0.027	0.063***	0.094***	0.094***	0.023*	0.029+	0.027+	0.002	0.002	0.001
	(0.019)	(0.020)	(0.020)	(0.010)	(0.015)	(0.015)	(0.011)	(0.016)	(0.016)	(0.003)	(0.005)	(0.005)
Work training	0.018	0.010	0.006	0.010	-0.004	-0.007	0.015	0.023**	0.022*	-0.004	-0.004	-0.003
	(0.015)	(0.012)	(0.013)	(0.007)	(0.006)	(0.007)	(0.009)	(0.008)	(0.009)	(0.004)	(0.004)	(0.004)
External Accountant	0.042**	0.034*	0.037*	0.020**	0.026**	0.026**	0.089***	0.079***	0.076***	0.014*	0.014*	0.015*
	(0.013)	(0.014)	(0.015)	(0.007)	(0.009)	(0.009)	(0.019)	(0.018)	(0.018)	(0.006)	(0.006)	(0.006)
Audit Employees	0.006	0.008*	0.010*	0.006**	0.006**	0.007***	-0.004+	-0.004	-0.002	0.001	0.001	0.001
	(0.003)	(0.004)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)
Time Sheet	0.002	0.000	0.000	0.002	0.000	-0.001	0.002	-0.001	0.001	-0.000	0.000	0.000
	(0.004)	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.000)	(0.001)	(0.001)
Terms w/benefits	0.015***	0.016***	0.014***	0.010***	0.008**	0.007**	0.003+	0.004*	0.004+	0.001	0.001	0.001
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Conflict Employees	-0.068**	-0.059*	-0.076**	-0.053*	-0.029	-0.027	-0.017	-0.010	-0.017	0.004*	0.004	0.009+
	(0.025)	(0.024)	(0.024)	(0.023)	(0.022)	(0.023)	(0.022)	(0.023)	(0.029)	(0.002)	(0.005)	(0.006)
Constant	0.642***			0.820***			0.843***			0.977***		
	(0.096)			(0.033)			(0.072)			(0.013)		
Observations	1897	1872	1864	1897	1872	1864	1897	1872	1864	1897	1872	1864
R-squared	0.162	0.263	0.289	0.105	0.337	0.353	0.042	0.187	0.220	0.016	0.087	0.098

Note: Estimated coefficients from LASSO model runs on Tax deduction (1-3); Payroll accounts (4-6); Monthly reporting (7-9) and general accounts (10-12). Standard errors in parentheses. Column (1, 4, 7 and 10) is OLS without fixed effects, column (2, 5 and 11) is OLS with fixed effects on NTA auditor, and column (3, 6 and 12) is OLS with fixed effects on NTA auditor and Nace code (2-digit). Residence time >10 years is the reference category and therefore omitted in the table. Standard errors are clustered by Nace code. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

Table A10: Results from the OLS Fixed Effects models on CPI Score and Conflict

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CPI Score	-0.000		0.000	-0.000		-0.000	-0.001		-0.001+
	(0.000)		(0.001)	(0.000)		(0.000)	(0.000)		(0.000)
Conflict		-0.031	-0.047		-0.007	0.003		-0.026	0.003
		(0.027)	(0.043)		(0.029)	(0.030)		(0.027)	(0.025)
Constant	0.896***	0.890***	0.886***						
	(0.012)	(0.007)	(0.014)						
Observations	1959	1974	1959	1936	1951	1936	1928	1943	1928
R-squared	0.000	0.001	0.001	0.139	0.135	0.139	0.211	0.206	0.211

Note: Estimated coefficients model runs on Compliance from CPI Score (columns 1, 4 and 7); Conflict (columns 2, 5 and 8) and both (columns 3, 6 and 9). Standard errors in parentheses. Column (1, 2 and 3) is OLS without fixed effects, column (4, 5 and 6) is OLS with fixed effects on NTA auditor, and column (7, 8 and 9) is OLS with fixed effects on NTA auditor and Nace code (2-digit). Standard errors are clustered by Nace code. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

Table A11: Fixed effects results from OLS, GA, and LASSO model runs

	(1)	(2)	(3)
Female	0.014		
	(0.021)		
Age	-0.003***		
	(0.000)		
Foreign	-0.050*		
	(0.023)		
<5	-0.165+		
	(0.095)		
5-10	-0.039		
	(0.036)		
Conflict	0.056*		
	(0.027)		
CPI Score	-0.000		
	(0.000)		
External Accountant	0.092***	0.111***	0.113***
	(0.023)	(0.021)	(0.023)
Conflict Employees	-0.152*	-0.119*	-0.124*
	(0.066)	(0.048)	(0.047)
Ltd Company		0.103***	0.145
		(0.022)	(0.103)
Salary system		0.103***	0.102***
		(0.029)	(0.029)
Work training		0.024	0.024
		(0.017)	(0.017)
Job advertisement		0.019+	
		(0.010)	
Self-employed			0.050
			(0.107)
Audit Employees			0.010*
			(0.004)
Time Sheet		0.003	0.001
		(0.007)	(0.006)
Terms		0.021***	0.020***
		(0.003)	(0.003)
Observations	1928	1864	1864
R-squared	0.236	0.313	0.314

Note: Estimated coefficients from Fixed effects on NTA auditor and Nace code (2-digit) model runs. Standard errors in parentheses. Column (1) is OLS, column (2) is GA, and column (3) is LASSO. Residence time >10 years is the reference category and therefore omitted in the table. Standard errors are clustered by Nace code. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10.

8. References

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A Theoretical Analysis. *Journal of Public Economics*, *1*(3-4), 323-338.

Alm, J. (2019a). Tax compliance and administration. In *Handbook on taxation* (pp. 741-768). Routledge.

Alm, J. (2019b). What motivates tax compliance? *Journal of economic surveys*, *33*(2), 353-388.

Alm, J., Deskins, J. A., & McKee, M. (2006). Third-party income reporting and income tax compliance. *Andrew Young School of Policy Studies Research Paper Series*(06-35).

Alm, J., Martinez-Vazquez, J., & McClellan, C. (2016). Corruption and firm tax evasion. *Journal of Economic Behaviour & Organization*, *124*, 146-163.

Alm, J., McClelland, G. H., & Schulze, W. D. (1992). Why do people pay taxes? *Journal of Public Economics*, *48*(1), 21-38.

Andreoni, J., Erard, B., & Feinstein, J. (1998). Tax compliance. *Journal of economic literature*, *36*(2), 818-860.

Ashby, J. S., Webley, P., & Haslam, A. S. (2009). The role of occupational taxpaying cultures in taxpaying behaviour and attitudes. *Journal of Economic Psychology*, *30*(2), 216-227.

Atawodi, O. W., & Ojeka, S. A. (2012). Factors that affect tax compliance among small and medium enterprises (SMEs) in North Central Nigeria. *International journal of business and management*, *7*(12), 87.

Bastani, S., Giebe, T., & Miao, C. (2020). Ethnicity and tax filing behaviour. *Journal of Urban Economics*, *116*, 103215.

Battaglini, M., Guiso, L., Lacava, C., Miller, D. L., & Patacchini, E. (2022). *Refining Public Policies with Machine Learning: The Case of Tax Auditing*.

Baum, A., & Gupta, S. (2017). Corruption, taxes and compliance. *eJTR*, *15*, 190.

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, *80*(6), 2369-2429.

Bellows, J., & Miguel, E. (2006). War and institutions: New evidence from Sierra Leone. *American Economic Review*, *96*(2), 394-399.

Benjamini, Y., & Hochberg, Y. (1995). 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.

Bird, R., & Davis-Nozemack, K. (2018). Tax avoidance as a sustainability problem. *Journal of business ethics*, *151*, 1009-1025.

Bjørneby, M., Alstadsæter, A., & Telle, K. (2018). Collusive Tax Evasion by Employers and Employees: Evidence from a Randomized Field Experiment in Norway.

Bjørneby, M., Alstadsæter, A., & Telle, K. (2021). Limits to third-party reporting: Evidence from a randomized field experiment in Norway. *Journal of Public Economics*, *203*, 104512.

Blackford, J. U. (2017). Leveraging statistical methods to improve validity and reproducibility of research findings. *JAMA psychiatry*, *74*(2), 119-120.

Bobek, D. D., Hageman, A. M., & Kelliher, C. F. (2013). Analyzing the role of social norms in tax compliance behaviour. *Journal of business ethics*, *115*, 451-468.

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, *8*, 3-62.

Boning, W. C., Guyton, J., Hodge, R., & Slemrod, J. (2020). Heard it through the grapevine: The direct and network effects of a tax enforcement field experiment on firms. *Journal of Public Economics*, *190*, 104261.

Bornman, M. (2015). The determinants and measurement of trust in tax authorities as a factor influencing tax compliance behaviour. *Journal of Economic and Financial Sciences*, *8*(3), 773-790.

Braithwaite, V., & Ahmed, E. (2005). A threat to tax morale: The case of Australian higher education policy. *Journal of Economic Psychology*, *26*(4), 523-540.

Broadhurst, D., Goodacre, R., Jones, A., Rowland, J. J., & Kell, D. B. (1997). Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry. *Analytica Chimica Acta*, *348*(1-3), 71-86.

Bø, E. E., Slemrod, J., & Thoresen, T. O. (2015). Taxes on the internet: Deterrence effects of public disclosure. *American Economic Journal: Economic Policy*, *7*(1), 36-62.

Cateni, S., Colla, V., & Vannucci, M. (2010). Variable selection through genetic algorithms for classification purposes. Proceedings of the 10th IASTED International Conference on Artificial Intelligence and Applications, AIA,

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., & Mullainathan, S. (2016). Productivity and selection of human capital with machine learning. *American Economic Review*, *106*(5), 124-127.

Clarke, D., Romano, J. P., & Wolf, M. (2019). The romano-wolf multiple hypothesis correction in stata.

Collier, P., & Sambanis, N. (2002). Understanding civil war: A new agenda. *Journal of Conflict Resolution*, *46*(1), 3-12.

Cummings, R. G., Martinez-Vazquez, J., McKee, M., & Torgler, B. (2009). Tax morale affects tax compliance: Evidence from surveys and an artefactual field experiment. *Journal of Economic Behaviour & Organization*, *70*(3), 447-457.

Damania, R., Fredriksson, P. G., & Mani, M. (2004). The persistence of corruption and regulatory compliance failures: theory and evidence. *Public choice*, *121*(3-4), 363-390.

DeBacker, J., Heim, B. T., & Tran, A. (2015). Importing corruption culture from overseas: Evidence from corporate tax evasion in the United States. *Journal of Financial Economics*, *117*(1), 122-138.

Del Giudice, V., De Paola, P., & Forte, F. (2017). Using genetic algorithms for real estate appraisals. *Buildings*, *7*(2), 31.

Dubin, J. A., & Wilde, L. L. (1988). An empirical analysis of federal income tax auditing and compliance. *National Tax Journal*, 61-74.

Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives*, *3*(4), 99-117.

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

Feld, L. P., & Frey, B. S. (2002). Trust breeds trust: How taxpayers are treated. *Economics of governance*, *3*(2), 87-99.

Feld, L. P., & Frey, B. S. (2007). Tax compliance as the result of a psychological tax contract: The role of incentives and responsive regulation. *Law & Policy*, *29*(1), 102-120.

Feldman, N., & Slemrod, J. (2009). War and taxation: When does patriotism overcome the free-rider impulse? *The new fiscal sociology: Taxation in comparative and historical perspective*, 138-154.

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

Fisman, R., & Miguel, E. J. J. o. P. e. (2007). Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. *115*(6), 1020-1048.

Foner, N. (2014). The Immigrant Family: Cultural Legacies and Cultural Changes1. In *The New Immigrant and the American Family* (pp. 111-124). Routledge.

Fonti, V., & Belitser, E. (2017). Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*, 1-25.

Frey, B. S., & Torgler, B. (2007). Tax morale and conditional cooperation. *Journal of comparative economics*, *35*(1), 136-159.

Gangl, K., Torgler, B., & Kirchler, E. (2016). Patriotism's impact on cooperation with the state: an experimental study on tax compliance. *Political psychology*, *37*(6), 867-881.

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M., & Strand, H. (2002). Armed conflict 1946-2001: A new dataset. *Journal of peace research*, *39*(5), 615-637.

Hanno, D. M., & Violette, G. R. (1996). An analysis of moral and social influences on taxpayer behaviour. *Behavioural research in Accounting*, *8*, 57-75.

Hansen, C., & Liao, Y. (2019). The factor-lasso and k-step bootstrap approach for inference in high-dimensional economic applications. *Econometric Theory*, *35*(3), 465-509.

Harding, L., & Mutascu, M. (2016). *Does migration affect tax revenue in Europe?*

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251-1271.

Hofmann, E., Hoelzl, E., & Kirchler, E. (2008). Preconditions of voluntary tax compliance: Knowledge and evaluation of taxation, norms, fairness, and motivation to cooperate. *Zeitschrift für Psychologie/Journal of Psychology*, *216*(4), 209-217.

Hofmann, E., Voracek, M., Bock, C., & Kirchler, E. (2017). Tax compliance across sociodemographic categories: Meta-analyses of survey studies in 111 countries. *Journal of Economic Psychology*, *62*, 63-71.

Hofstede, G. (1984). *Culture's consequences: International differences in work-related values* (Vol. 5). sage.

Jahnke, B., & Weisser, R. A. (2019). How does petty corruption affect tax morale in Sub-Saharan Africa? *European Journal of Political Economy*, *60*, 101751.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Jenkins, D. G., & Quintana-Ascencio, P. F. (2020). A solution to minimum sample size for regressions. *PloS one*, *15*(2), e0229345.

Joulfaian, D. (2000). Corporate income tax evasion and managerial preferences. *Review of Economics and Statistics*, *82*(4), 698-701.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2005). *Governance matters IV: governance indicators for 1996-2004*. The World Bank.

Kirchler, E. (2007). *The economic psychology of tax behaviour*. Cambridge University Press.

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. *79*(3), 651-692. https://doi.org/10.3982/ecta9113

Kleven, H. J., Landais, C., Saez, E., & Schultz, E. (2014). Migration and wage effects of taxing top earners: Evidence from the foreigners' tax scheme in Denmark. *The Quarterly Journal of Economics*, *129*(1), 333-378.

Kopczuk, W., Marion, J., Muehlegger, E., & Slemrod, J. (2016). Does tax-collection invariance hold? evasion and the pass-through of state diesel taxes. *American Economic Journal: Economic Policy*, *8*(2), 251-286.

Kountouris, Y., & Remoundou, K. (2013). Is there a cultural component in tax morale? Evidence from immigrants in Europe. *Journal of Economic Behaviour & Organization*, *96*, 104-119.

Lange, T., & Melsom, A. M. (2022). Påvirker oppholdstid i Norge etterlevelse hos utenlandske arbeidstakere? *Samfunnsøkonomen*, *1*, 28-39.

Lassen, D. D. (2007). Ethnic divisions, trust, and the size of the informal sector. *Journal of Economic Behaviour & Organization*, *63*(3), 423-438.

Liu, H.-H., & Ong, C.-S. (2008). Variable selection in clustering for marketing segmentation using genetic algorithms. *Expert Systems with Applications*, *34*(1), 502-510.

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

Marandu, E. E., Mbekomize, C. J., & Ifezue, A. N. (2015). Determinants of tax compliance: A review of factors and conceptualizations. *International Journal of Economics and Finance*, *7*(9), 207-218.

Miguel, E., Saiegh, S. M., & Satyanath, S. (2011). Civil war exposure and violence. *Economics & Politics*, *23*(1), 59-73.

Muehlbacher, S., Kirchler, E., & Schwarzenberger, H. (2011). Voluntary versus enforced tax compliance: Empirical evidence for the "slippery slope" framework. *European Journal of Law and Economics*, *32*(1), 89-97.

Mullainathan, S., & Spiess, J. J. J. o. E. P. (2017). Machine learning: an applied econometric approach. *31*(2), 87-106.

Musimenta, D., Nkundabanyanga, S. K., Muhwezi, M., Akankunda, B., & Nalukenge, I. (2017). Tax compliance of small and medium enterprises: a developing country perspective. *Journal of Financial Regulation and Compliance*.

Nordblom, K., & Žamac, J. (2012). Endogenous Norm Formation Over the Life Cycle-The Case of Tax Morale. *Economic Analysis & Policy*, *42*(2).

Norwegian Tax Administration, T. (2019). *Kunnskapskontroller på lønnsområdet*

Teknisk rapport.

Pearl, J. (2018). Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv preprint arXiv:1801.04016*.

Pereira, J. M., Basto, M., & da Silva, A. F. (2016). The logistic lasso and ridge regression in predicting corporate failure. *Procedia Economics and Finance*, *39*, 634-641.

Potocky, M., & Naseh, M. (2020). *Best practices for social work with refugees and immigrants*. Columbia University Press.

Romano, J. P., & Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, *100*(469), 94-108.
Sable, S., Porwal, A., & Singh, U. (2017). Stock price prediction using genetic algorithms and evolution strategies. 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA),
Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, *81*(10.5555), 26853.
Sandmo, A. (2005). The theory of tax evasion: A retrospective view. *National Tax Journal*, *58*(4), 643-663.
Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, *18*(5), 429-434.
Sefiane, S., & Benbouziane, M. (2012). Portfolio selection using genetic algorithm.
Slemrod, J. (2019). Tax compliance and enforcement. *Journal of economic literature*, *57*(4), 904-954.
Snow, A., & Warren, R. S. (2005). Ambiguity about audit probability, tax compliance, and taxpayer welfare. *Economic Inquiry*, *43*(4), 865-871.
Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human resources*, *50*(2), 301-316.
Statistics Norway, T. (2022). www.ssb.no
Saad, N. (2014). Tax knowledge, tax complexity and tax compliance: Taxpayers' view. *Procedia-Social and Behavioural Sciences*, *109*, 1069-1075.
Tittle, C. R. (1980). Sanctions and social deviance: The question of deterrence.
Tolvi, J. (2004). Genetic algorithms for outlier detection and variable selection in linear regression models. *soft computing*, *8*(8), 527-533.
Torgler, B. (2004). *Tax morale, trust and corruption: Empirical evidence from transition countries*.
Torgler, B. (2007). *Tax compliance and tax morale: A theoretical and empirical analysis*. Edward Elgar Publishing.
Torgler, B., Demir, I. C., Macintyre, A., & Schaffner, M. (2008). Causes and consequences of tax morale: An empirical investigation. *Economic Analysis and Policy*, *38*(2), 313-339.
Transparency International. (2019). *Corruption Perceptions Index 2018: Score timeseries since 2012*, . https://www.transparency.org/en/cpi#
Tsakumis, G. T., Curatola, A. P., & Porcano, T. M. (2007). The relation between national cultural dimensions and tax evasion. *Journal of international accounting, auditing and taxation*, *16*(2), 131-147.

Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research*, *14*(Mar), 867-897.

Watson, L. (2015). Corporate social responsibility, tax avoidance, and earnings performance. *The Journal of the American Taxation Association*, *37*(2), 1-21.

Witte, A. D., & Woodbury, D. F. (1985). The effect of tax laws and tax administration on tax compliance: The case of the US individual income tax. *National Tax Journal*, 1-13.

Young, C., & Varner, C. (2011). Millionaire migration and state taxation of top incomes: Evidence from a natural experiment. *National Tax Journal*, *64*(2), 255.

Zhou, Z., Zhang, Q., & So, A. M.-C. (2015). p-Norm Regularization: Error Bounds and Convergence Rate Analysis of First-Order Methods. International Conference on Machine Learning.

Hard vs Soft Intervention: Compliance Effects among Firms[8]

Thomas Lange

Abstract

Audits and letters are two of the enforcement strategies available to a tax administration to ensure compliance. In this paper we use a unique set of experiments to determine which of the two enforcement strategies is more effective. We use firm-level data from 1,974 randomized audits and 8,000 information letters. We find that audits cause an immediate and significant increase in the firm's remittance of payroll tax, but less strong effects from letters. Updated, perceived audit probability seems to sustain adjustments in payroll tax remittance two years post-treatment. Firms receiving the information letters also adjust their remittance upwards, and more so when the letters are actually read. Our 'back of the envelope' cost-benefit calculations suggests that tax administrations could save resources by partially switching to cheaper enforcement strategies, like information letters.

JEL Classification: H2; H25; H26; H32; C31; C36

[8] Thanks to Pablo Garriga, Evelina Gavrilova-Zoutman, Andreas Kotsadam, Knut Løyland, Anne May Melsom, Hanne Beate Næringsrud, Nina Serdarevic, Floris Zoutman, and Arnstein Øvrum for useful insights and comments.

1. Introduction

Firms play a central part in the economy and account for a significant fraction of the collected tax revenue. In 2021, the tax revenue ratio of firms/individual taxpayers in Norway was .75, not including VAT (Norwegian Ministry of Finance, 2022). The payroll tax is a significant contribution to the Norwegian tax revenue. In 2021, Norwegian firms paid NOK 216.4 billion in total payroll tax, equalling 14.2 percent of the total tax revenue, or 22.9 percent of taxes liable to firms only (Norwegian Ministry of Finance, 2022). Employers are responsible for the remittance and the reporting of the payroll tax on behalf of their employees, as part of the financing of the National Insurance Scheme. The tax is levied on employees' salaries and other taxable remuneration for work and assignments in and outside of an employment relationship. The tax rates are set by the Norwegian Parliament every year and range from 5.1 percent to 14.1 percent depending on the sector and geographical location.

Firms' tax behaviour and exposure to the tax administration are different from those of the individual taxpayers. While the latter is obligated to submit their tax return only once a year, a firm has several reporting and payment obligations throughout the year, such as the bi-monthly reporting of the payroll tax remittance, and the monthly submission of the VAT return. Thus, the higher frequency of encounters with the tax authorities may affect the firms' compliance in different ways than individual taxpayer compliance, requiring different enforcement strategies. Since audits are generally costly, whereas standardized information letters are cheap, tax administrations may save considerable resources by moving from audits to letters.

This paper contributes to recent literature on the effects of different tax enforcement strategies. Our experimental setting captures central aspects of the real-world reporting environment such as the presence of firms' bi-monthly remittance of the payroll tax. We compare the effects on firms' remittance of payroll tax of a scalable strategy, namely standardized information letters with the effects of on-site audits. Unlike most other contributions, we study the effects of these two instruments on the same population, and in the same institutional setting. The attention is directed to a population of 30,961 firms in different labour-intensive businesses in the Norwegian economy.

Since our two treatments (on-site audits and standardized, electronic information letters) are directed towards firm's reporting obligations, we seek to uncover whether the firm respond to and correct their payment of payroll tax, after treatment with an audit or a letter stating the required obligations. We measure compliance through remitted amounts of payroll tax. We believe payroll tax is of particular relevance in this context, both because it is a tax only levied on firms, and the message conveyed in

the treatments is targeted towards employee reporting obligations, which indirectly influence payroll tax remittance.

The audits were implemented on a stratified random sample of 1,974 firms during the filing and auditing season of 2018. The audited firms were given an up-front notification about the content of the audit, so that the firms could prepare relevant documentation. The letters were sent to the same population, but to a different sample of 8,000 recipients simultaneously in June 2018. A fraction of 14 percent (1,130) of the letter recipients never opened the letter and was therefore not treated.

Audits and information letters represent two enforcement strategies, with different expected effects on a firm's behaviour, and at very different administrative costs for the tax authorities. On-site audits are more intrusive than impersonal information letters. While a scheduled visit from a tax auditor typically evokes a sense of being subject to the disclosure of potential errors, a letter stating the firm's reporting liabilities is perceived more as a reminder of the applicable rules than a sense of being inspected. One may therefore expect that a change in behaviour succeeding an audit originates in a sense of enforcement, whereas a prospective change following an information letter may be perceived as more 'voluntary.' On the Tax administration's end, an audit requires time-consuming actions, depending on the nature and scope of the audit, whereas a letter, once written, is practically cost-free to distribute. A change in enforcement strategy could therefore result in significant cost savings for the Tax Administration. The main question is then: What kind of compliance effects can we expect from the two treatments?

The empirical analysis is divided into two main parts. The first part studies the average treatment effects (hereafter ATE) of audits and letters on firms' remittance of payroll tax. As we are interested in a comparison with Bjørneby et al. (2021) on employee effects, we keep the number of employees as a dependent variable alongside the analysis of the payroll tax. While on-site audits are more comprehensive in that they involve more scrutiny than standardized letters, we expect considerable variation in remitted payroll tax across firms depending on the assigned treatment.

We find that the audit treatment increased the firm's payroll tax remittance in 2018 by 13.20 percent compared to the pre-treatment level in 2017, and we find a statistically significant increase in remitted payroll tax in all post-treatment years (2018-2020) for the audited firms compared to the reference group receiving no treatment. Furthermore, we find a similar but less strong effect of 1.82 percent for the firms receiving a letter. This confirms the findings of Ortega and Scartascini (2020) who find stronger effects from physical visits than emails. The effect seems to be more stable for the audited firms, two years following the treatment (2020). This is consistent with

previous findings such as Kleven et al. (2011), DeBacker et al. (2018) and Advani et al. (2018).

The second part of the analysis studies the Local Average Treatment Effects of the letters (hereafter LATE).[9] This part allows us to isolate the effects of the firms that actually read the letters, i.e. the treated (or 'compliers') in this experiment, from the whole group of firms that one intended to treat (ITT). The nature of the letter treatment differs from that of audits in this respect; there is often one-sided non-compliance in that a fraction of firms receiving a digital letter will leave it unread. Thus, we study the effects of being treated, using an IV model capturing this effect. We find that the compliers remit significantly higher payroll taxes in all subsequent years following the treatment. This is consistent with the findings in Bjørneby et al. (2021).

The policy implications of the results in this paper are two-fold. First, our analysis reveals that there is a significant compliance effect of alternative enforcement strategies like information letters. Second, since information letters are scalable to a larger population at low cost, such enforcement may be preferable to traditional, more expensive on-site audits. We include in this part a limited cost-benefit analysis, indicating that letters may be preferable in the short run.

The paper is organized as follows. Section 2 reviews adjacent literature, Section 3 describes the institutional background and the random audit program. Section 4 describes the data. Section 5 lays out the experimental design. Section 6 estimates the effect of the two treatments, Section 7 suggests some policy implications, and Section 8 concludes.

2. Related literature

The rational agent-based theory of Allingham and Sandmo (1972) has shaped the modern tax administration's approach to enforcement by increasing taxpayer's perceived costs of evasion, decreasing the cost of compliance, and tailoring enforcement strategies towards different taxpayer segments (Baer & Silvani, 1997). More recently, behavioural and moral aspects of tax compliance have also found their way into tax research (e.g. Dhami and Al-Nowaihi (2007) and Luttmer and Singhal (2014), and have gained increased attention among tax administrations. Rather than increasing audit frequency, modern tax administrations have sought to gain knowledge of compliance effects from other, prospectively cheaper and more scalable enforcement strategies (Murphy (2019); Alm (2019a); Keen and Slemrod (2017)). Tax administrations may save scarce audit resources by switching to more effective enforcement strategies.

[9] Sometimes referred to as Complier Average Causal Effect (CACE) (Gerber and Green 2012).

The compliance effects of enforcement strategies in general, and audits in particular, have recently been studied in experimental designs, using randomized samples (Kotsadam et al. (2021); Kleven et al. (2011); Advani et al. (2018); DeBacker et al. (2018); Hebous et al. (2020); Bjørneby et al. (2021)). Most of these studies use samples of individual taxpayers except Almunia and Lopez-Rodriguez (2018), which uses a non-random sample of firms from Large Taxpayers Unit (LTU) in Spain, Boning et al. (2020) which uses total employer tax deposits, and Bjørneby et al. (2021) which uses on-site audits of Norwegian firms from a stratified sample. Unlike these contributions which study the effects of enforcement on the individual taxpayer, we study the effects of audits and letters as enforcements on the firm. The novelty of our contribution is the study of how the firm as an entity may respond in a very different way than individual taxpayers. Furthermore, we study the effects of two enforcements on the payroll tax – a tax to which only the firm is liable, in the same environment. D'Agosto et al. (2018) also studied the compliance effect on small businesses of two different enforcement strategies, namely on-site and desk-based audits. However, they use risk-based audits, not random selection. Our approach also differs from random designs like that of Boning et al. (2020) who studies the effects on overall employment taxes, i.e. payroll taxes and employee income taxes withheld and remitted by IRS-assigned at-risk firms. Since the latter can be manipulated by the employee, even in a third-party remittance regime (Bjørneby et al., 2021), part of the effects in Boning et al. (2020) may be explained by employee non-compliance or collusive actions involving both employer and employee. We overcome these issues by focusing on the payroll tax alone.

The effects of audits on compliance vary across studies, sensitive to which dependent variable is chosen, but Kleven et al. (2011), DeBacker et al. (2018) and Advani et al. (2018) find lasting effects on subsequent tax compliance among audited taxpayers. These effects are confirmed by Løyland et al. (2019) using risk-based audits, and Hebous et al. (2020). Because audits are costly, tax administrations also use cheaper and less intrusive enforcement policies such as information campaigns, reminders, enforcement emails (Brockmeyer et al., 2019), letters (Pomeranz (2015); Doerrenberg and Schmitz (2015)) and encouragements (Kotsadam et al., 2021). The effects of such soft treatments are also mixed, depending on the message portrayed in the treatment (Alm (2019b); Slemrod (2019); Meiselman (2018); Pomeranz and Vila-Belda (2019)). Information that increases perceived detection probability seems to have positive short-term effects on reporting (Blumenthal et al. (2001); Kleven et al. (2011); Fellner et al. (2013); Bott et al. (2020)), while general appeals to tax morale and social norms seem to have little or no effects (Hallsworth et al., 2017), or even negative effects (De Neve et al., 2021). Using four different letter treatments, Bergolo et al. (2023) find that information on audits *decreased* the perceived probability of being audited, while, at the same time, inducing a significant deterrent effect on tax evasion.

Our design is similar to that of Boning et al. (2020), except that the 12,172 firms they study are ex-ante suspected of non-compliance, selected by algorithm based on payments before and during the fourth quarter of 2014, i.e. firms showing signs of noncompliance before treatment. While it is highly useful for tax administrations to acquire knowledge about treatment effects on risk firms, our contribution differs in that our random selection is not limited to a population of firms suspected of non-compliance, but rather to a fraction of the economy, irrespective of previous risk-based selection. Our findings are representative of a population of 30,961 firms from labour-intensive businesses. Our approach provides a broader view of the compliance effects of the treatments within the entire sector. Boning et al. (2020) is a more targeted approach, allowing for confirmation or rejection of the initial suspicions, and providing evidence of different behaviour among high-risk firms. In their setting, a fraction of the firms also received an information letter, and another fraction had an on-site IRS Revenue Officer visit.

While the unit of study in Kotsadam et al. (2021) is a sample of individual taxpayers claiming deductions, we direct the attention towards 30,961 firms in different labour-intensive businesses in the Norwegian economy. Taxpayers claiming deductions are more prone to errors than the rest of the taxpayerpopulation, and so the external validity of the effects observed in Kotsadam et al. (2021) may not extend to a population of firms. What motivates individual taxpayer behaviour may also differ from the determinants behind firm compliance.

Bjørneby et al. (2021) study a compliance mechanism evolving from third-party reporting, using randomized audits of firms in Norway, and find compliance effects of the audit treatment. We may expect similar effects from the audits in our study, but we move beyond their setup and compare the effects from the audits with prospective effects from letters. This may give tax administrations information about more cost-efficient enforcement strategies.

3. Institutional Background

3.1 The Random Audit Program

In 2017, the Norwegian Tax Administration (NTA) introduced the random audit program to build more systematic knowledge about tax compliance. The program is conducted along three thematic strands: labour market regulations, VAT compliance, and quality of third-party data reporting. We focus on data from the first strand.

The audits were directed towards disclosing the scope and magnitude of formal compliance errors in labour-intensive businesses in Norway. This focus originated from risk-based audits, where experience indicates that formal non-compliance is particularly high where foreign labour is involved. However, risk-based audits are

biased, and so the inference from the audit sample to the general population is limited. This random audit program therefore sought to measure the compliance gap in labour-intensive industries to make more robust conclusions about compliance risk in the population of these industries. The audits had reporting items on e.g. firms' accounting- and salary systems, recruitment routines, board and lodging for foreign employees, use of foreign workers and subcontractors, tax withholding accounts, salary reporting and staff registers, if applicable.

The program was thus established to gain knowledge of reporting non-compliance rather than enforcing tax remittance. However, correct reporting will inevitably entail correct tax remittance, but rather than disclosing evaded taxes, the program aimed at disclosing errors in the firms' reporting procedures.

The random audit program started with a test pilot in 2017. A total of 60 test audits were performed in this initial phase. These audits are not included in the data we use in this paper. Following an evaluation of this pilot, adjustments were made before the start of the program. In its full scope, 187 auditors from the NTA have been involved in executing the audits. Each audit averaged three to five days of work, with an average cost of 12,740 NOK per audit. All 1,974 audits were executed in 2018.

3.2 The Information Letters

The letter recipients were randomly drawn from the same target population as the audited firms. The information letters contained descriptive information on 7 relevant reporting duties for the firms in these sectors but contained no 'moral' statements as in e.g. Bott et al. (2020). While the audits were more extensive than the letters, the reporting duties stated in the letters reflected some of the same items on which the audited firms were asked to provide, namely accounting standards, monthly reporting (A-melding), employee tax deductions, documentation of salary expenses, documentation of elapsed time, reporting duties to the International Tax Collection Office, and staff register obligations. A copy of the letter can be found in Appendix 3. The marginal cost of electronically distributing the letters, and the average cost per letter are both negligible.

3.3 The target population and sample selection

The target population for both the audits and the letters is based on the selection criteria in Table 1 and consists of 30,961 firms.

Table 1. Population selection criteria

Criterion
1. Total revenue in 2017 > NOK 100,000
2. If not 1, then total revenue in 2016 ≥ NOK 200,000.
3. Mean active work relations per month is ≤ 20
4. Mean active employees per month is ≥ 5
5. If not 4, then registered > NOK 100,000 on subcontracts and/or foreign services in 2016
6. Private Limited Companies (AS) self-employed (ENK) and Norwegian foreign-registered enterprises (NUF)
7. No termination date registered
8. Sectors in public service, defence and public social security are excluded

Note: The table describes the selection criteria for firms in the Strata described in Appendix Table A2, which resulted in the 30,961 firms in the population of study. The selection criteria are set to secure representative firm activity (criteria 1-2; 7), employment (3-5) organisational form (6), and exclude the public sector (8).

The 8,000 firms receiving letters were randomly drawn from this target population using simple random selection and no stratification or any further selection criteria. With this increased sample size, compared to the audits, the likelihood of obtaining a representative and diverse sample that reflects the overall population characteristics increases, and hence stratification becomes less important (Solon et al., 2015). No audited firms received a letter, however. Thus, no firms received two treatments, and there was no interference between treatment groups. The electronic letters were sent out simultaneously to all 8,000 firms on 1 June 2018, of which 6,870 firms did open/read the letters. The NTA can read off which firms opened the letters, and which firms left them unopened in the digital government dialogue ("Altinn").

Out of 2,000 randomly selected firms, 26 were found "unworthy" of an audit for different, unsystematic reasons. The 1,974 audited firms were selected through a stratified random sample to capture the relative sizes of the different strata in the total population of 30,961 firms. A total number of 22 strata representing different industries were constructed from 65 different NACE codes on 1- and 2-digit levels, cf. Appendix 1. The sample size allocated to each stratum was determined by the method of proportional allocation (based on the number of foreign employees in each industry). The samples were then selected from each industry (stratum) for each of the five tax regions in Norway separately, considering the audit resources capacity in each region. The aim was at least five audits in each stratum in each region, but the number of audits in each stratum could not exceed 15 percent of the total resource capacity in the region (proportional allocation method with lower and upper cut-off). A list of the industry sector and stratum is included in the Appendix.

4 Data description

4.1 Descriptive Statistics

The data set covers the period 2017-2020. While the shortcoming of this limited period is of some concern for establishing post-treatment trends, it is evident from other studies that the strongest effects occur immediately or within a couple of years post-treatment (e.g. Advani et al. (2021); Brockmeyer et al. (2019); Pomeranz (2015)), DeBacker et al. (2018) being an exception, finding stronger effects on total income in years 3 and 4 compared to years 1 and 2 post audit. Our data set covers a wider period than e.g. Boning et al. (2020) and Brockmeyer et al. (2019), and an additional post-treatment year compared to Bjørneby et al. (2021).

To provide an overview of the different sectors' representation in the sample, the descriptive statistics are broken down by stratum, as displayed in Appendix Table A1, while Table A2 displays which sectors are included in the strata. Construction of buildings and civil engineering (S26), Human health, residential care, and social work (S15), Publishing, broadcasting, telecommunications and Information Services (S17), food and beverage service activities (S31) and Manufacturing (S16) are by far the largest strata in the population concerning the number of firms in each stratum.

Payroll tax represents a significant tax revenue contribution from the firms in our population, where the total reported contributions vary from NOK 8.5 to 9.8 billion over the years of study. Table 2 shows reported remittance of payroll tax in NOK, and firms' workforce, broken down by population, reference, audit and letter groups for the years 2017-2020.

Table 2. Reported remittance of Payroll Tax (NOK) and number of employees (2017-2020).

		Payroll Tax			Employees		
	Year	N	Mean	SD	N	Mean	SD
Population	2017	30 869	300 651	416 841	30961	11.95	11.41
	2018	30 813	311 326	550 902	30961	11.99	14.25
	2019	30 807	318 893	702 543	30961	11.46	17.6
	2020	30 667	285 473	733 289	30961	10.22	15.6
Reference	2017	20 919	300 638	423 045	20987	11.72	11.07
	2018	20 893	310 461	586 372	20987	11.71	13.68
	2019	20 884	319 539	776 792	20987	11.15	16.84
	2020	20 786	287 767	815 942	20987	9.99	16.06
Audit Arm	2017	1 970	284 822	342 069	1974	14.05	14.42
	2018	1 964	299 351	388 336	1974	14.21	14.92
	2019	1 966	299 264	434 835	1974	13.49	15.67
	2020	1 964	253 812	417 955	1974	11.82	14.24
Letter Arm	2017	7 980	304 594	417 213	8000	12.04	11.42
	2018	7 956	316 552	485 016	8000	12.18	15.47
	2019	7 957	322 046	529 603	8000	11.78	19.82
	2020	7 917	287 302	539 802	8000	10.41	14.63

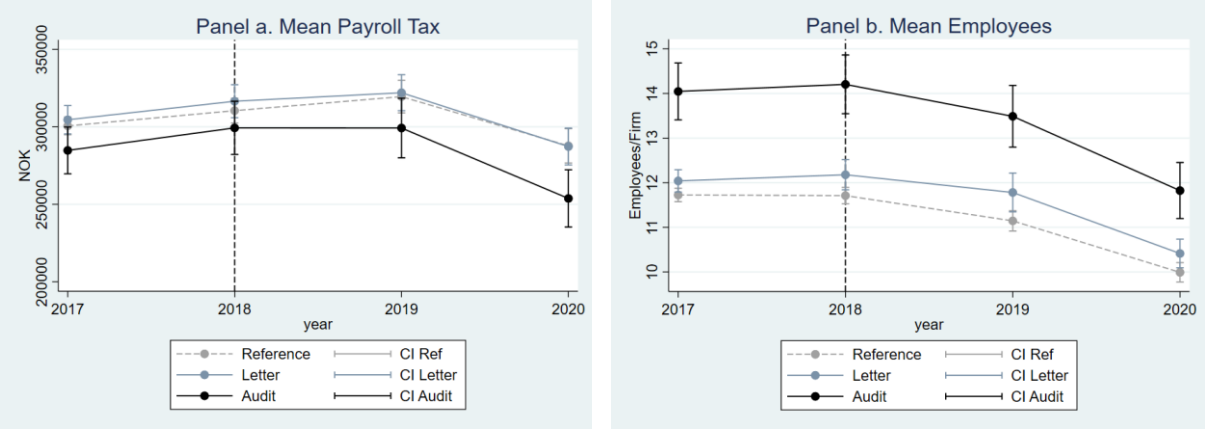
Note: Reported remittance of payroll tax in NOK. Column 'N' denotes total number of observations/firms in the respective years given by the first column. Column 'Mean' represents sample means, and column 'SD' gives the standard deviation. 735 erroneous, negative values are omitted. Random checks on negative values indicate no evidence that omitted observations are systematic.

The letter group is four times as large as the audit group, and so we expect lower standard errors on the estimated effects of letters than of the audits. The reference group is the same population of firms for both audit and letter arms. The standard deviation is high for all groups, and similar across groups. All groups' payroll tax remittances increased from the pre-treatment year 2017 to the treatment year 2018. As expected for a true experimental design, we found no evidence of the Ashenfelter dip when adding a pre-treatment year (2016) in the context of any interventions wherein those assigned to either treatment have a temporarily depressed payroll tax remittance that would revert upward toward their longer-term mean, absent treatment (Ashenfelter, 1978), cf. Appendix Figure AF1. As expected, due to randomization, the treatment groups are similar before treatment, except that the audit group had a systematically higher number of employees compared to all other groups during the whole period of study. This is due to the stratification of the audit sample, where overrepresented strata had firms with higher numbers of employees, cf. t-test statistics with strata fixed effects in Table 3.

4.2 Graphical evidence

A simple way to test the effect of a treatment is to track the means of the outcome variables, pre- and post-treatment, for the treatment and control groups respectively, cf. Figure 1.

Figure 1. Means of Payroll tax and number of employees. Reference, Audit and Letter groups



Notes: This figure plots point means and 95% confidence intervals from payroll tax remittance (in NOK), and the number of employees by years 2017 to 2020. The treatment, either audits or letters, take place in year 2018, visualised by the vertical, dotted line. The black line represents the annual mean of the audit group, the blue line represents the annual mean of the letter group, and the dotted line represents the mean of the reference group. The 95% confidence intervals are represented by the vertical error bars for each group.

From 2018 to 2019, all three groups slightly increased their payroll tax remittance compared to the pre-treatment year of 2017, and then decrease towards 2020. Thus, on average, we cannot infer that the treatments affect firms' payment of payroll tax. The firms' workforce is also downward sloping from 2017 onwards, with a sharper decline from the treatment year of 2018 for all three groups. On the average, we do not see any effect of treatment on the firms' workforce.

As we can infer from both panels, randomization fails on raw data because there are significant differences between the groups, pre-treatment. However, since the audit sample was stratified, we need to control for strata fixed effects when testing randomization.

4.3 Proof of Randomization

To test whether the treatment groups and reference group were significantly different in the pre-treatment year of 2017, we ran a simple OLS regression with fixed effects on strata for the audit arm to control for the stratification bias, but no fixed effects for the letter arm because the letter sample selection was not stratified:

(1) $Y_{ist} = \alpha_i + \beta_t Audit_{is}^{*} 2017 + \varepsilon_{ist}$
(2) $Y_{it} = \alpha_i + \gamma_t Letter_i^{*} 2017 + \varepsilon_{it}$

Y_{ist} is the log of payroll tax for firm *i* at time *t* = 2017 in stratum *s*. β_t and γ_t are unbiased and consistent estimators for the audit and letter treatments in year 2017, α_i and η_t are firm constants and ε is the error term.

Furthermore, there are reasons for clustering the standard errors, either a sampling design reason, because we have sampled data from a *population* using clustered sampling (Liang & Zeger, 1993), and an experimental design reason (Weiss et al., 2016), because the assignment mechanism for the audit treatment is clustered. Albeit both reasons are relevant for this study, at this stage of the analysis, we use clustered standard errors adopting the first reason. The test results are displayed in Table 3.

Table 3. T-statistics Payroll tax and Employees in the base year 2017

Arm	Payroll Tax	t	P>|t|	Employees	t	P>|t|
Audit	3658.342	0.54	0.592	0.297	0.92	0.366
	(6721.756)			(0.321)		
Letter	5317.589	1.60	0.124	0.116	0.61	0.548
	(3322.063)			(0.190)		
Observations	30869			30961		
R-squared Audit	0.055			0.124		
R-squared Letter	0.000			0.000		

Note: OLS estimation coefficients and test statistics for the audit arm compared to the stratified audit reference arm, and letter arm compared to the unstratified letter reference arm. Estimated coefficients on Y = Payroll Tax and Y = Employees. Fixed effects on strata for audit arm. Standard errors are adjusted for 22 clusters in strata. Robustness test is run with standard error clusters on firm ID, and the result stands.

As we can see from the T-statistics in Table 3, randomization holds for both audit and letter arms on payroll tax and Employees.

4.4 Pre-treatment comparison

Because we are also interested in a comparison of treatment effects on the two treatment arms, we have run a two-sample t-test with equal variances on payroll tax and Employees to reveal prospective pre-treatment deviances between the treatment arms on the two variables. The test results are displayed in Table 4.

Table 4. Two sample T-statistics Payroll tax and Employees in the base year 2017

Arm	Payroll Tax	t	P>|t|	Employees	t	P>|t|
Audit	284821.6			14.07107		
	(342068.8)			(14.42238)		
Letter	304594			12.05564		
	(417212.9)			(11.42717)		
Combined	300679.2	1.9480	0.0514	12.45467	-6.6322	0.0
	(403508.8)			(12.10517)		

Note: Test statistics for Audit (N=1,970) and Letter (N=7,980) arms on Y = Payroll Tax and Y = Employees. 24 Erroneous, negative values are omitted. Standard deviations in parenthesis.

The audit and letter arms are statistically significant different on Employees, but not on Payroll Tax. The explanation for this difference is the audit selection criterion of mean active employees per month ≥ 5. The letter sample selection had no such criterion.

5 Experimental design

In a randomized experiment, where treatments are assigned to test and control groups by random selection, an OLS model with fixed effects would give unbiased estimates. To exploit the panel structure and to adjust for the fact that the two groups have different average levels of compliance in the base year 2017, we use a difference-in-difference model design (hereafter DiD). However, the results are equivalent to an OLS. There are two reasons for this. First, a DiD increases the precision of the estimates (Angrist & Pischke, 2009). Second, we want to estimate LATE for the letter treatment group. As a robustness check, we have run simple OLS fixed effects models. The results are displayed in Appendix Table A3.

The two treatment samples have different properties in three relevant respects. Firstly, the audit sample is stratified, and so we fix the effects on strata for this sample. Secondly, all firms assigned to audit treatment were actually audited. Thus, for the audit treatment group, treatment assignment is identical to the treatment status, and so ITT estimations give us the ATE. Lastly, the letter sample is not stratified but includes non-compliers, i.e. a known sub-sample of those who received the letter but did not open it. Therefore, they are untreated but nevertheless assigned to the letter treatment sample. We estimate both ATE for this group and LATE for those who were actually treated.

The letter recipients and the audited firms are drawn from the same target population, using simple random sampling for the letters and stratified random sampling for the audits. To ensure that the treatment effects are comparable we include fixed effects on

strata in models including both treatments. This allows for comparison of treatment effects within each stratum.

5.1 ATE of treatments

While evaded tax or changes in reported income tax are commonly used dependent variables (Kotsadam et al. (2021); Bjørneby et al. (2021)), we are interested in changes in payroll tax as an expression of a correction following the treatments. As less reported payroll leads to more profits, more dividends and then an increase in income, the economic incentive for withholding is present.

Our first model utilises payroll tax as the dependent variable, with fixed effects on strata, firm ID and year, to isolate the effects from the treatment:

(3) $Y_{ist} = \alpha_i + \eta_t + \beta_t Audit^*_{is} Year + \gamma_t Letter^*_{is} Year + \varepsilon_{ist}$

Y_{ist} is the log of payroll tax for firm i at time t = 2017…2020 in stratum s. β_t and γ_t are unbiased and consistent estimators for the audit and letter treatments, interacted by years 2017-2020, α_i and η_t are firm, and time constants, respectively, and ε is the error term. We interact the treatments with year and so we estimate the effects of each year, compared with pre-treatment year t = 2017. We also run a model where Y_{ist} is the log of the workforce for firm i at time t = 2017…2020 in stratum s. Standard errors are clustered on strata. The number of clusters is only 22, and so we may underestimate the true standard errors (Angrist & Pischke, 2009). Since a two-sided Wald test would be a poor approximation, we have run the regressions with standard errors clustered on firm ID (n=30,961) and NACE code at 2-digit level (n=61) as well, without notable deviations, cf. Appendix Table A5.

To allow for a comparison between ATE and LATE, we run separate models on the annual effects of treatments compared to the reference group. For the audit arm (4) we include fixed effects on strata and firm ID, and for the letter arm (5) we include fixed effects on firm ID:

(4) $Y_{st} = \beta_t Audit_s + \varepsilon_{st}$
(5) $Y_t = \gamma_t Letter + \varepsilon_t$

Y is the log of payroll tax at time t = 2017…2020 in stratum s. β_t and γ_t are unbiased and consistent estimators for the audit and letter treatments, respectively in the years 2017-2020, and ε is the error term. Again, we interact the treatments with year and so we estimate the effects of each year, compared with pre-treatment year t = 2017. As for (3), we also run (4) where Y_{st} is the log of the workforce at time t = 2017…2020 in stratum s. Standard errors are clustered on strata.

5.2 LATE of letter treatment

There is a difference in effects between the group one intended to treat with the letter and the effects of the group in fact being treated, i.e. the group opening/reading the letter. The former estimates the effect of being assigned to a treatment group, whereas the latter estimates the effect of being treated. This difference arises from the fact that there may be firms assigned to treatment, that end up not getting treated.

LATE is thus an estimate that focuses on a specific subgroup within the population, i.e., the compliers. LATE estimates the average effect of the treatment on this subgroup, which is typically smaller than the overall population. It is based on the assumption that there are no unmeasured confounding variables affecting treatment assignment for the compliers. In other words, it assumes that the treatment effect is constant for this subgroup, regardless of the level of treatment received by others. ATE, on the other hand, aims to estimate the average effect of the treatment on the entire population, including both compliers and non-compliers. It considers the effect of treatment on all individuals, regardless of whether they fully comply, partially comply, or do not comply with the treatment. For the audits, we have no such differences, since all firms assigned to the audit treatment actually did get audited.

For the letter treatment group, there was one-sided non-compliance, i.e. firms who received but didn't open the letters, but no firms who didn't receive the letters and still required/read them. Some of the firms assigned to letter treatment received the letter, but never opened them. However, no firms assigned to the reference group received/read the letter. The ATE gives the effects for the whole group intended to get the letters, but there are no reasons to expect effects among firms who never opened the letter. To estimate the effect only among firms who were actually treated, called the Locale Average Treatment Effect (LATE), we use the IV-model described in the following two equations:

(6) $Letter_i = \alpha + \beta Treat_i + u_{is}$
(7) $Y_{it} = \eta + \gamma \widehat{Letter_i} Year_{it} + \varepsilon_i$

$Treat_i$ is a dummy which takes the value 1 if firm i was assigned to the letter treatment group, controlled for time, t whereas $Letter_i$ indicates whether the firm actually opened/read the letter, a is a constant, β is the letter treatment estimator, and u is the error term. Standard errors are clustered in strata. We estimate the parameters β and γ through a Two-Stage Least Squares (2SLS) with (6) as the first stage and (7) as the second stage.

Letter assignment is the instrument variable. It affects treatment as only letter recipients may read the letters. However, it is not associated with the dependent

variables (payroll tax and number of employees) as firms are randomly assigned to treatment.

Given that the assignment to the letter group was true random on payroll tax, our reduced form ATE estimates (4) and (5) can be given causal interpretation of the ATE. The IV estimates (6) and (7) rely on two additional assumptions (Angrist and Pischke (2009); Gerber and Green (2012)).

First, the non-interference assumption, which consists of two parts. Part A stipulates that whether a firm is treated depends only on the firm's own treatment group assignment. Because there is no firm receiving a letter outside the letter treatment group, this condition holds. Part B stipulates that potential effects are affected by the firm's own assignment, *and* the treatment the firm receives as a consequence of that assignment (Gerber & Green, 2012, p. 138).

Second, the exclusion restriction stipulates that potential effects respond to actual treatments, not treatment assignments. That is, firms respond to the letter only if they open/read it, not by merely receiving it. There are good reasons why we should assume that the exclusion restriction holds because treatment does not affect the outcome variables through any other channels than through the direct reading of the letter. Since communication between the NTA and firms in Altinn is not unusual, it is unlikely that firms might respond by merely receiving a letter from the NTA.

6 Results

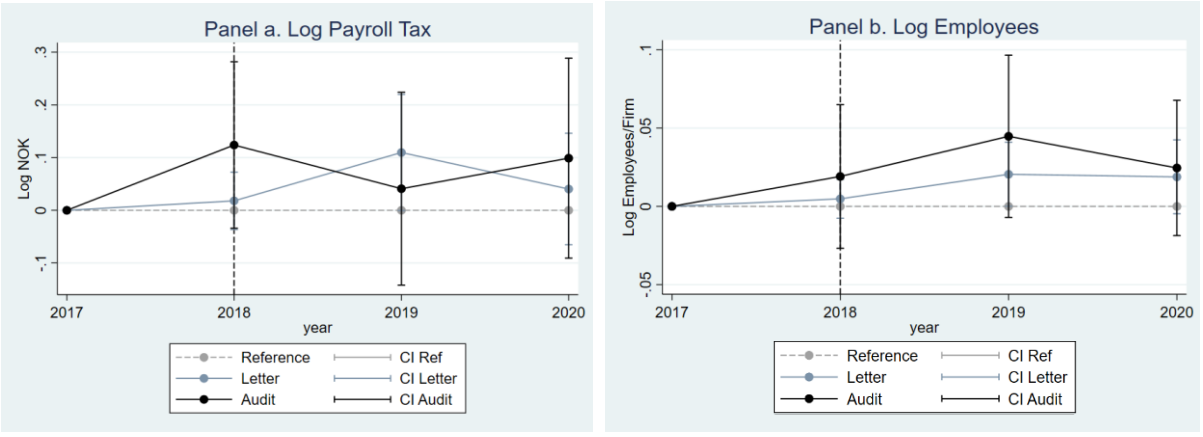
6.1 Main results

The two treatments may affect a firm's reporting liabilities other than the payroll tax. Revenue, salary expenses and number of employees (workforce) may be affected by both the audit and the letter treatments since these are specific items contained in both treatments. We have run all the models in this paper with these independent variables as well, but we find no major deviations from the main model using payroll tax remittance as the dependent variable. Audits may affect payroll tax remittance through both better documentation on the payroll tax itself, but also through better documentation of sales and other variables. We are estimating the total effect of all these possible channels, cf. correlation matrix in Appendix Table A4. Nevertheless, as we are interested in a comparison with Bjørneby et al. (2021) on employee effects, we keep the number of employees as a dependent variable alongside the analysis of the payroll tax.

The regression results from our combined model (3) are presented in Figure 2. The results reveal a positive effect of audit and letters compared to the reference group

when we use the pre-treatment year of 2017 as the base year, but the audit and letter curves have opposite shapes post-treatment.

Figure 2. Estimated Audit and Letter (ATE) Effects (base year 2017)



Notes: This figure plots point DiD estimates and 95 percent confidence intervals from regressions of measures of logged payroll tax remittance and logged number of employees for the years 2017-2020, compared to the base year 2017, and compared to the reference group. (Difference between treatment and reference in years 2018; 2019; 2020, minus the difference between treatment and reference in 2017). The treatment, either audits or letters, take place in year 2018, visualised by the vertical, dotted line. The black line represents the estimated coefficients of the audit group, the blue line represents the estimated coefficients of the letter group, both estimations compared to the reference group represented by the dotted line. The specification includes fixed effects on strata and firm ID. Standard errors are adjusted for 22 clusters in strata. Table A5 in the Appendix displays the coefficients and their standard errors.

The audit treatment had an immediate effect in 2018 on firms' payroll tax remittance compared to the reference group. In 2018, the audited firms increased payroll tax remittance by .124 log points (equivalent to 13.20 percent), compared to the reference group when we use the pre-treatment base year 2017.

Increased remittance following an audit may be considered "mechanical" when firms only adjust their behaviour and report accurate information in response to the specific discrepancies or errors uncovered during the audit. The increased remittance may thus not be driven by a genuine compliance commitment or improved understanding of reporting liabilities, but rather a reaction to the fear of penalties or consequences for non-compliance. This identification problem pertains to the distinction between actual behavioural changes in taxpayers (real effects) and mere adjustments made to comply with tax liabilities in response to the audit (reporting effects) (Advani et al. (2021); Kleven et al. (2011)). Real effects refer to substantive changes in the behaviour of taxpayers resulting from the audit. These changes may include a genuine improvement in tax compliance practices, a better understanding of tax regulations, and a shift towards more accurate and honest reporting of financial information

(Kausar et al., 2016). Reporting effects, on the other hand, occur solely in response to the audit itself. Firms may correct errors or discrepancies discovered during the audit, but these changes might not reflect any meaningful change in their overall tax compliance behaviour beyond the specific issues identified during the audit. Reporting effects are often temporary, and taxpayers may revert to non-compliant behaviour once the audit process is over. The immediate effect in 2018 and the slight dip in 2019 may indicate reporting effects, but the following increase in 2020 makes this difficult to establish.

The effect of letters on payroll tax remittance is .018 log points (equivalent to 1.82 percent) in 2018. The audited firms also increased their workforce in 2018 by .019 log points (equivalent to 1.92 percent), compared to the reference group, and the letter group increased their workforce by .005 (equivalent to .50 percent) in the treatment year 2018. We see the strongest effects on payroll tax remittance of audits in 2018, decreasing in 2019, and then increasing in 2020. The letter treatment follows an opposite pattern, where we find the strongest effects in 2019, and then decreasing in 2020. In 2019, the effect of letters on payroll tax remittance is .110 log points (equivalent to 11.63 percent), which is closer to the 2018-estimate for the audited firms. Thus, there appears to be a lag in the letter treatment effect, not apparent in the audit treatment.

The time-lagged compliance effect from the information letter, as compared to the immediate effect from the on-site audits, can be attributed to the different nature of these two compliance interventions and how they influence taxpayer behaviour over time. The letters are non-binding and serve as educational tools to inform firms about their compliance obligations. Unlike the audits, the letters do not involve direct enforcement actions or a perceived threat of penalties for non-compliance. As a result, firms may take longer to internalize the information and voluntarily adjust their behaviour. Behavioural change takes time, regardless of the method used to convey information. After receiving an information letter, firms may need time to process the information, assess its implications, and gradually adopt better practices.

An on-site audit typically leads to more immediate compliance adjustments due to the direct scrutiny and enforcement actions involved. The fear of penalties and the immediate presence of auditors can compel firms to address identified compliance issues more promptly.

The treatment effect on the workforce is consistent with the findings in Bjørneby et al. (2021) who estimate an average increase of 1.12 in the number of employees between treated and untreated firms. The stronger effect from the letter treatment compared to the audit treatment on payroll tax remittance in 2019, can be seen in light of a similar finding in D'Agosto et al. (2018), which reveals a higher effect from the 'soft' on-site

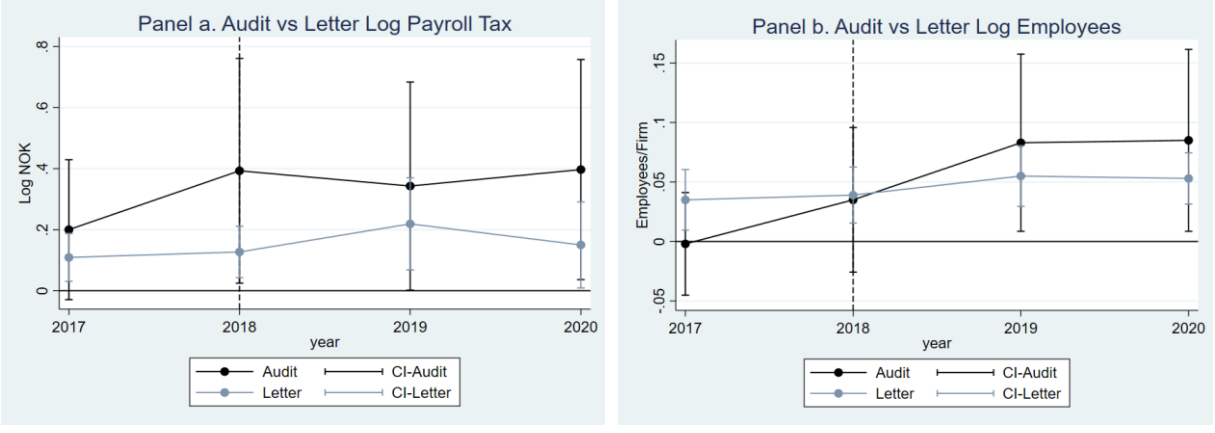
audit compared to the 'deep' desk-based audit treatments in their study. But as our treatments differ in nature, i.e., our audit treatment resembles their on-site audits rather than their desk-based audits, a direct comparison should be considered with caution.

The letter treatment has significant effects on both outcome variables in the first post-treatment year of 2019. Letters affect payroll tax and workforce less than audits two years post-treatment. This confirms the findings in Boning et al. (2020), where letters conveying the same message as on-site visits by the Revenue Officers in the U.S. have smaller direct effects. The same pattern is evident in Ortega and Scartascini (2020) who find stronger effects from on-site visits than emails. Although Kotsadam et al. (2021) find a similar response of letter treatment (diminishing with time), the results are not directly comparable; their unit of study is the individual, not the firm, and the letters are more specifically addressing the actual source of error, namely deductions on the tax return. Doerrenberg and Schmitz (2015) also suggest evidence that a letter which reminds small firms of the civic duty to pay taxes and informs about an audit probability following the letter, may increase tax compliance, but their results are not statistically significant.

6.2 Annual Audit and Letter effects

To allow for a comparison between ATE and LATE, we have run separate models (4) and (5). The results are displayed in Figure 3.

Figure 3. Estimated Annual Audit and Letter (ATE) Effects 2017-2020



Notes: This figure plots point estimates and 95 percent confidence intervals from annual regressions (run separately for audit and letter group) of measures of logged payroll tax remittance and logged number of employees for the years 2017-2020, compared to the reference group (x-line). The treatment, either audits or letters, take place in year 2018, visualised by the vertical, dotted line. The black line represents the estimated coefficients of the audit group, the blue line represents the estimated coefficients of the letter group, both estimations compared to the reference group. The specification includes strata fixed effects (audits). Standard errors are adjusted for 22 clusters in strata. Table A6 in the Appendix displays the coefficients and their standard errors.

The audit effect on payroll tax exceeds that of the letters by a substantial margin. An audit would increase the average firm's remittance of payroll tax by .393 log points more than the non-treated firms in the treatment year 2018 (equivalent to 48.14 percent). Correspondingly, a letter would increase the average firm's payroll tax remittance by .127 log points more than the non-treated firms in this year (equivalent to 13.54 percent). Whereas the audit effect sustains two years post-treatment (2020), the letter effect is stronger one-year post-treatment (2019) than two years post-treatment (2020). The audit effect on firms' workforce is more ambiguous but appears also to exceed that of the letter group in the post-treatment years 2019-2020. An audit would increase the workforce by .083 and .085 log points compared to the reference group in the post-treatment years 2019 and 2020, respectively (equivalent to 8.65 and 8.87 percent), whereas the letter effect is .039 and .055 respectively (equivalent to 3.98 and 5.65 percent).

6.3 Heterogeneous Effects

In the second chapter of this thesis, we found significant differences in compliance effects between firm types. To study prospective Heterogeneous treatment effects, we have run all models on samples restricted to the three firm types represented in the population, namely Private limited liability company (Private Ltd), self-employed and Norwegian registered foreign company. By and large, Private Ltd's appear to drive the main results. The significance levels of estimated coefficients are higher on this sub-sample, which is over 80 percent of the firms in the population, cf. Appendix Table A9. This heterogeneity resembles the results of Almunia and Lopez-Rodriguez (2018) who also find that the impact of monitoring on tax compliance varies across different types of firms. Larger and more profitable firms are more responsive to increased monitoring, whereas smaller and less profitable firms show a weaker response. Brockmeyer et al. (2019) find an ambiguous Heterogeneous effect between corporations and the self-employed in that the filing rate of corporations responds less strongly but their payment rate responds more strongly to the treatment compared to the self-employed. Pomeranz (2015) finds a stronger response on VAT compliance for smaller firms, following a letter informing about a random audit selection among 400,000 Chilean firms.

We find some small significant negative treatment effects among the self-employed, cf. Appendix Table A10. This is contrary to the findings of Mittone et al. (2017). They suggest that tax audits have a significant short-term impact on increasing compliance. Their treatment group tend to immediately adjust their behaviour and report more accurately to address the specific issues identified during the audit. However, the study also reveals a gradual decline in tax compliance levels after the initial surge

following the audit. Over time, the fear of audit diminishes, and taxpayers may return to their previous, less compliant behaviour or adopt tax avoidance strategies to reduce the risk of future audits, known as "the bomb-crater effect" (Mittone et al., 2017). As our significance levels are low, and not reproduced for the letter treatment, this result should be treated with caution. Beer et al. (2020) find both pro-deterrent and counter-deterrent effects on future reporting behaviour of audits among self-employed. They suggest that the observed reduction in reported income among self-employed U.S. taxpayers may be associated with dishonesty caused by non-detected misreporting during the audit. According to their study, such taxpayers may infer that audits are ineffective, and cause self-employed to understate their income even more aggressively in subsequent years. Since the sub-sample of self-employed is limited in our experiment, and the significant negative treatment effect is found on the reported number of employees, not on payroll tax remittance, we cannot infer that the mechanism suggested by Beer et al. (2020) is at play in our environment.

We find no significant effects among Norwegian registered foreign companies (cf. Appendix Table A11), which is also expected since this company type is less than 1 percent of the total sample.

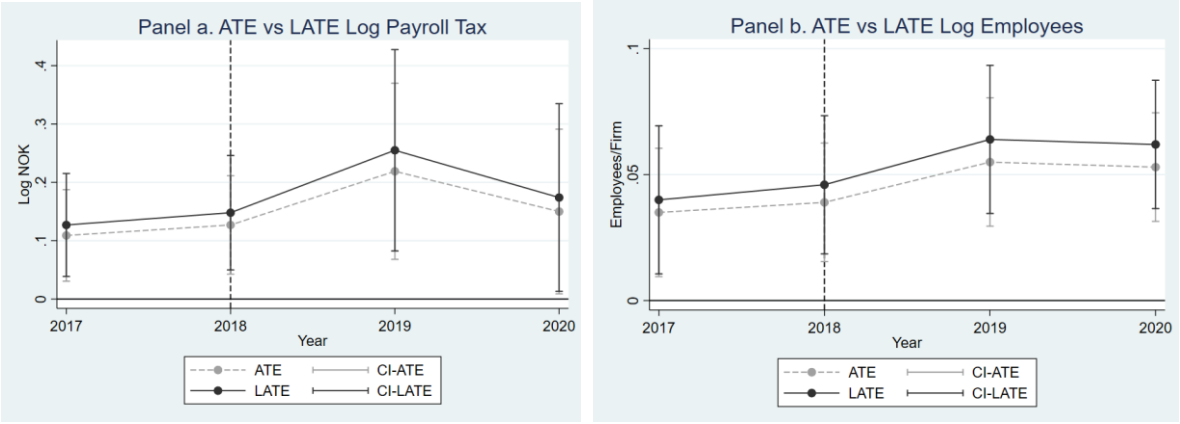
We have also run all models on samples restricted to firm size in the lower and upper 25 percentiles on firm revenue. We find no significant effects on payroll tax nor employees from either treatment on firms in the lower and upper 25 percentiles measured by firm revenue, cf. Appendix Table A8. Using number of employees as a proxy for firm size, Bergolo et al. (2023) finds no substantial or statistically significant difference between firms below and above the median number of employees.

6.4 Effects of reading the letter (LATE)

To get a clearer picture of the letter treatment effects, we estimate the LATE to see if there is any difference between those who actually read the letters and those who did not. Of the 8,000 firms who received a letter, 1,130 never opened/read it. Thus, the ITT estimates presented above may be biased, since just above 14 percent of the letter treatment group was in fact never treated. While the ITT estimates measure the effects of receiving a letter, the LATE estimates measure the effects of actually reading the letter. This is equivalent to the set-up in Bjørneby et al. (2021), except that we deal only with one-sided non-compliance, as there were no firms in the reference group receiving a letter.

The regression results, presented in Figure 4 reveal a positive effect of reading the letters (compliers) compared to the reference group.

Figure 4. Estimated LATE of letter treatment (base year 2017)



Notes: This figure plots annual point IV-estimates and 95 percent confidence intervals from regressions of measures of logged payroll tax remittance and logged number of employees for the years 2017-2020, compared to the reference group (x-line). The letter treatment took place in year 2018, visualised by the vertical, dotted line. The black line represents the estimated coefficients of the 6,870 letter openers (LATE), the blue line represents the estimated coefficients of the 8,000 assigned to the letter group. Standard errors are adjusted for 22 clusters in strata. Table A7 in the Appendix displays the coefficients and their standard errors.

Both in the treatment year 2018, and the post-treatment year 2019, we observe that the compliers report significantly higher payroll tax (.148 and .255 log points, equalling 16.0 and 29.0 percent, respectively), and number of employees (.046 and .064 log points equalling 4.7 and 6.6 percent, respectively) compared to the reference group. The effects seem to remain two years post-treatment (2020). Furthermore, the statistically significant estimates may indicate that being assigned to the letter treatment is a strong instrument of actually being treated.

The significant difference between the LATE and the ATE estimates tells us that there may be further gains by taking low-cost measures to facilitate letter reading, such as an automated reminder.

7 Policy implications

These findings may inform resource allocation decisions. We study only the direct effects on payroll tax remittance and the workforce of the firms and assume few or small network effects as studied by e.g. Boning et al. (2020).

Any treatment would increase net revenue if the marginal revenue it raises exceeds its marginal administrative costs. In this paper, we limit the revenue component to payroll tax only, even if there may be other benefits, such as network effects, and a general deterrent effect in the overall population of firms. The revenue raised should

be compared to the marginal administrative cost of the treatments. The equation which must be true for implementing either treatment is thus:

(8) $R_A - C_A > 0$
(9) $R_L - C_L > 0$

R is revenue (payroll tax) from either two treatments, audit or letter denoted by subscripts A and L, respectively, and C is the marginal, administrative cost. The "back of the envelope" calculations are based on the estimated effects from models (4) and (5) in 2018 and are displayed in Table 5.

Table 5. Cost-benefit analysis

	Audit (A)	Letter (L)	Reference (R)	A-R	L-R
Marginal Cost (MC)	12 740	0	0	12 740	0
β (%) 2018	0.4814	0.1354	0	0.4814	0.1354
Δ (NOK) 2017-2018	14 529	11 958	9 823	4 706	2 135
β*Δ = Marginal Revenue (MR)	6 994	1 619	0	2 265	289
MR-MC	-5746	1 619	0	-1 075	289

Note: Marginal cost of audits is NOK 12,740. β is the estimated coefficient (%) in year 2018 from model (2) and (3) respectively. Δ is the increment in NOK from 2017-2018. A-R gives the difference in NOK between Audit and Reference, and L-R gives the difference in NOK between Letter and Reference.

From Table 4, we see that an audit will increase payroll tax remittance by 48.14 percent more than the reference group, whereas the corresponding figure for the letter treatment is 13.54 percent. Thus, 48.14 and 13.54 percent of the increment is due to the treatment. On average, a firm in the audit group increased its payroll tax remittance from 2017 to 2018 by NOK 14,529, whereas the figure for a firm in the letter group was NOK 11,958. By comparison, a firm in the reference group increased its payroll tax remittance by NOK 9,823. Thus, the difference in payroll tax remittance between an audited firm and the reference group is NOK 4,706 on average, and the corresponding figure for a firm receiving a letter is NOK 2,135. If we apply the estimated coefficients to these figures, NOK 2,265 of the payroll tax increment is due to the audit and NOK 289 of the increment is due to the letter. Hence, the administrative cost of the audit exceeds the revenue it generates, and equation (8) is rejected, whereas the opposite is true for the letter treatment, and equation (9) holds.

Some caveats remain, however: First, it is too early to draw any conclusion on the long-term effects of both treatments, and previous literature suggests either declining effects over time (Boning et al., 2020), or mixed long-term effects (Bott et al., 2020). Second, there may also be other effects from both treatments, like general deterrence effects, which may increase the benefits. Third, even without such a deterrence effect, the

letters can still help guide the firms that mis-report due to honest mistakes, but the additional information will do very little to reduce tax evasion. Fourth, the rejection of (8) might also reflect that randomized audits are wasteful if the objective of those audits is increased compliance measured by revenue on the firm-level, which is, however, not the case. Randomized audits are used by tax administrations for objectives *other* than non-compliance disclosure on firm-level, like disclosing new areas of non-compliance or building datasets for predictive modelling (Alm (2019a); Micci-Barreca and Ramachandran (2004)). Finally, an electronic letter is easily scalable to a larger population, which, all else equal, will increase the revenue further.

8 Concluding remarks

There is a need to robustly evaluate different enforcement strategies' effectiveness to choose the most efficient and cost-effective enforcement strategy. Tax authorities may save scarce resources by switching from hard to soft interventions. The main contribution of this paper is a documentation of the firm's response to two interventions in an experimental setting involving the same population of firms. We utilize two randomized experiments, one with stratified on-site audits, and one using electronic letters. We demonstrate unbiased, positive effects on firms' remittance of payroll tax from both audits and letters, compared to a reference group with no treatment. While audits have stronger effects than letters, the former is by far the most expensive enforcement strategy.

The results are specific to a population of labour-intensive firms in a relatively advanced tax reporting environment, by international standards. Thus, the external validity of the results is not restricted to the Norwegian setting as such, as one can assume that these effects will be reproduced in several advanced, Western tax jurisdictions with gold standard system of information reporting. Furthermore, the results suggest that tax authorities may test and compare these two enforcement strategies in other sectors as well.

A cost-benefit assessment including estimations of deterrence effects and other prospective treatment benefits is necessary to make clear recommendations on which enforcement strategy to use. Such an assessment should still be a core priority of the modern tax administration, to increase the effects of limited enforcement resources.

9 Appendix

9.1 Population

Table A1. Descriptive statistics by stratum. Base year 2017.

Strata	N	Payroll tax (μ)	Employees (μ)	SD	Audit (n)	SD	Letter (n)	SD
S1	199	479 174	429 103	11	8	22	51	41
S11	505	110 167	270 168	13	10	34	113	89
S12	1 904	328 739	509 941	14	16	148	472	415
S13	1 097	256 028	173 401	11	6	46	297	261
S15	3 288	255 995	241 945	10	10	67	877	757
S16	2 372	399 841	523 591	12	9	89	637	567
S17	3 065	447 220	637 986	10	12	76	861	746
S19	277	96 385	140 913	17	22	124	47	35
S20	753	147 690	182 643	14	18	53	177	159
S21	1 701	323 216	310 702	11	9	91	425	359
S23	875	342 038	321 279	10	8	91	246	214
S26	5 484	301 787	347 375	9	8	300	1 395	1 211
S27	368	257 976	248 780	25	14	92	79	73
S29	621	243 176	238 360	14	10	132	122	106
S3	1 809	104 121	120 253	10	9	32	505	399
S31	2 815	233 333	197 782	23	13	302	681	564
S33	775	222 703	258 052	8	7	91	203	173
S34	948	493 216	997 972	11	9	48	261	230
S37	146	500 468	669 059	13	10	20	34	34
S38	1 359	342 466	305 586	10	7	77	359	313
S5	151	453 951	607 734	11	10	20	39	36
S6	449	246 548	201 366	11	5	19	119	88

Notes: Column 'Strata' reflects the strata description in Table A2. Column 'N' reflects the number of observations in each stratum. Column 'Payroll tax (μ)' reflects the sample means of Payroll tax (NOK) in each stratum, and column 'Employees (μ)' reflects the sample mean number of employees in each stratum. Columns 'Audit (n)' and 'Letter (n)' reflect the number of firms in each stratum receiving audit and letter in treatment year 2018. Column 'SD' is Standard Deviation.

Table A2. Stratum and Nace codes

Stratum	NACE 2	Description	N
S1	0	Missing NACE	31
S1	43	Specialised construction activities	144
S1	71	Architectural and engineering activities; technical testing and analysis	21
S1	90	Creative arts and entertainment activities	3
S11	3	Fishing and aquaculture	505
S12	77	Rental and leasing activities	102
S12	78	Employment activities	407
S12	79	Travel agency tour operator and other reservation service and related activities	231
S12	80	Security and investigation activities	104
S12	81	Services to buildings and landscape activities	480
S12	82	Office administrative office support and other business support activities	580
S13	96	Other personal service activities	1097
S15	86	Human health activities	2544
S15	87	Residential care activities	28
S15	88	Social work activities without accommodation	716
S16	10	Manufacture of food products	204
S16	13	Manufacture of textiles	59
S16	14	Manufacture of wearing apparel	33
S16	15	Manufacture of leather and related products	6
S16	16	Manufacture of wood and of products of wood and cork except furniture; manufacture of articles of straw and plaiting materials	228
S16	18	Printing and reproduction of recorded media	255
S16	20	Manufacture of chemicals and chemical products	2
S16	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	5
S16	22	Manufacture of rubber and plastic products	8
S16	23	Manufacture of other non-metallic mineral products	60
S16	24	Manufacture of basic metals	29
S16	25	Manufacture of fabricated metal products, except machinery and equipment	591
S16	26	Manufacture of computer electronic and optical products	81

S16	27	Manufacture of electrical equipment	55
S16	28	Manufacture of machinery and equipment n.e.c.	152
S16	30	Manufacture of other transport equipment	27
S16	31	Manufacture of furniture	116
S16	32	Other manufacturing	195
S16	33	Repair and installation of machinery and equipment	266
S17	58	Publishing activities	640
S17	59	Motion picture video and television programme production sound recording and music publishing activities	421
S17	60	Programming and broadcasting activities	29
S17	61	Telecommunications	87
S17	62	Computer programming consultancy and related activities	1665
S17	63	Information service activities	223
S19	1	Crop and animal production hunting and related service activities	277
S20	90	Creative arts and entertainment activities	480
S20	91	Libraries archives museums and other cultural activities	14
S20	93	Sports activities and amusement and recreation activities	259
S21	49	Land transport and transport via pipelines	1701
S23	43	Specialised construction activities	875
S26	41	Construction of buildings	5182
S26	42	Civil engineering	302
S27	55	Accommodation	368
S29	81	Services to buildings and landscape activities	621
S3	49	Land transport and transport via pipelines	1809
S31	56	Food and beverage service activities	2815
S33	43	Specialised construction activities	775
S34	50	Water transport	376
S34	51	Air transport	30
S34	52	Warehousing and support activities for transportation	404
S34	53	Postal and courier activities	138
S37	37	Sewerage	45

S37	38	Waste collection treatment and disposal activities; materials recovery	97
S37	39	Remediation activities and other waste management services	4
S38	45	Wholesale and retail trade and repair of motor vehicles and motorcycles	1359
S5	6	Extraction of crude petroleum and natural gas	19
S5	7	Mining of metal ores	2
S5	8	Other mining and quarrying	55
S5	9	Mining support service activities	75
S6	47	Retail trade, except of motor vehicles and motorcycles	449

9.2 OLS Design fixed effects model results

Table A3. OLS Design fixed effects model results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Arm	2017	2018	2019	2020	2017	2018	2019	2020
Audit	0.198+	0.386*	0.337+	0.399*	0.000	0.034	0.079*	0.082*
	(0.114)	(0.185)	(0.168)	(0.179)	(0.022)	(0.031)	(0.037)	(0.038)
Letter	0.095*	0.114*	0.206*	0.136+	0.038**	0.043**	0.058***	0.055***
	(0.039)	(0.041)	(0.075)	(0.069)	(0.011)	(0.011)	(0.013)	(0.011)
Observations	30961	30961	30961	30961	30961	30961	30961	30961
R-squared	0.045	0.030	0.026	0.026	0.112	0.089	0.058	0.042

Note: Columns (1)-(4) are Log Payroll Tax, Columns (5)-(8) are Log Employees. Fixed effects on strata. Standard errors in parenthesis. Standard Errors are adjusted for 22 clusters in strata. The coefficients are to be interpreted in log points and compared to a reference group with no treatment. + p<0.10; * p<0.05; ** p<0.01; *** p<0.001.

9.3 Information Letter (English Translation)

Information on enterprises' reporting obligations

The Tax Administration is working on making it as simple as possible to report correctly. Below, you can read about the most important obligations you have when running a business. You can find more information on each of the obligations at http://www.skatteetaten.no/rapporteringsplikt.

Obligation to report

Accounting
Most enterprises must keep accounts according to Norwegian bookkeeping legislation.

A-melding
Enterprises with salaried employees have, among other things, an obligation to report information about the employee's employment. This can be information about when they start, any changes in their employment relationship and when the employment is ended. In addition, they must report salary and other remuneration in the a-melding, along with information about withholding tax and other deductions made upon payment. The a-melding must be submitted every month.

Tax deductions
Enterprises must make tax deductions from salary payments and retrieve their employees' tax deduction cards. Deducted tax must be deposited in a separate bank account by no later than the first working day after the salary payment.

Documentation of salary and other benefits subject to a reporting obligation
The documentation of salary and other benefits subject to a reporting obligation must show remunerations per employee. The following information must follow from the documentation: National identity number, name and position, tax municipality, table number and/or the deduction percentage as stated on the tax deduction card.

Documentation of accrued time
Enterprises performing services where the payment is based on accrued time must document completed hours. The hours must be specified for each employee per day. The documentation must be finished by the end of the month when the work has been performed.

Reporting to the Central Tax Office – Foreign Tax Affairs (SFU)
All assignments and any subcontracts that are given to foreign contractors must be reported to the Central Tax Office - Foreign Tax Affairs (SFU). The condition is that the assignment has been performed at a place for building assembly work in Norway, at a site that is under the control of the client in Norway or on the Norwegian continental shelf.

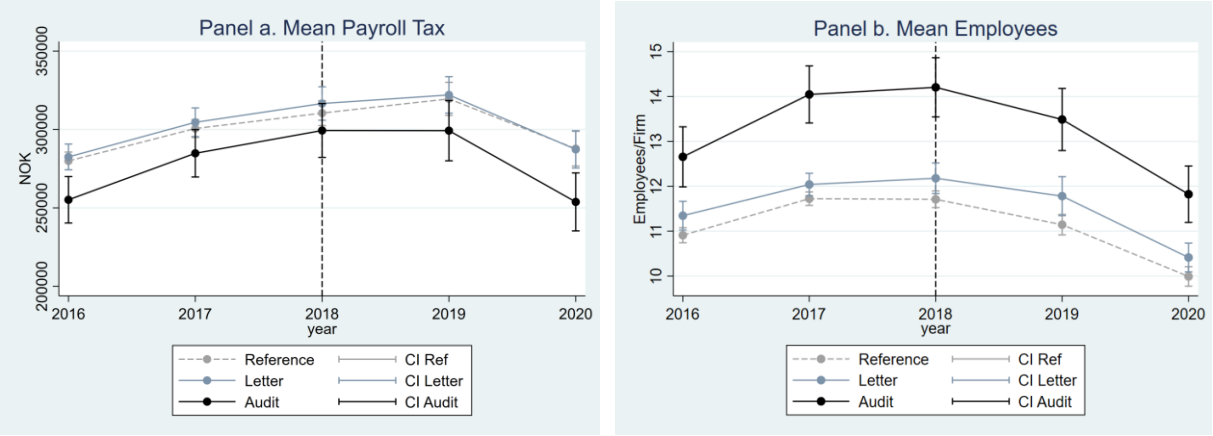
Staff register
Catering establishments, hairdressers, beauticians, car repair workshops and businesses in the car care sector must keep a staff register. A staff register is a list of everyone who works for the enterprise. The register must show when the employees start and finish their workday. The register must be kept at the workplace and be available for inspection during the opening hours of the business.

Yours sincerely

The Norwegian Tax Administration

9.4 Pre-Treatment Trend, 2016 included

Figure AF1. Means of Payroll tax and number of employees. Reference, Audit and Letter groups



Notes: This figure plots point means and 95% confidence intervals from payroll tax remittance (in NOK), and the number of employees by years 2016 to 2020. The treatment, either audits or letters, take place in year 2018, visualised by the vertical, dotted line. The black line represents the annual mean of the audit group, the blue line represents the annual mean of the letter group, and the dotted line represents the mean of the reference group. The 95% confidence intervals are represented by the vertical error bars for each group.

9.5 Correlation between dependent variables

Table A4. Correlation Matrix dependent variables

	Log Payroll tax	Log Revenue	Log Employees	Log Salary
Log Payroll tax	1.000			
Log Revenue	0.613	1.000		
Log Employees	0.495	0.395	1.000	
Log Salary	0.909	0.644	0.530	1.000

9.6 Regression results tables

Table A5. Estimated Audit and Letter (ATE) Effects (base year 2017)

Treatment	(1)	(2)	(3)	(4)	(5)	(6)
Audit*2018	0.124	0.124+	0.124	0.019	0.019+	0.019
	(0.081)	(0.068)	(0.079)	(0.023)	(0.012)	(0.022)
Audit*2019	0.041	0.041	0.041	0.045	0.045*	0.045
	(0.093)	(0.098)	(0.101)	(0.026)	(0.018)	(0.027)
Audit*2020	0.099	0.099	0.099	0.025	0.025	0.025
	(0.097)	(0.109)	(0.108)	(0.022)	(0.023)	(0.025)
Letter*2018	0.018	0.018	0.018	0.005	0.005	0.005
	(0.028)	(0.040)	(0.034)	(0.006)	(0.007)	(0.006)
Letter*2019	0.110+	0.110*	0.110*	0.021+	0.021+	0.021+
	(0.056)	(0.054)	(0.049)	(0.010)	(0.011)	(0.011)
Letter*2020	0.040	0.040	0.040	0.019	0.019	0.019
	(0.054)	(0.062)	(0.043)	(0.012)	(0.013)	(0.012)
Observations	123844	123844	123844	123844	123844	123844
R-squared	0.756	0.756	0.756	0.835	0.835	0.835

Note: Regressions with Fixed effects on strata and firm ID. The coefficients are to be interpreted in log-points. Standard errors in parenthesis. Columns (1) - (3) are Log Payroll Tax and Columns (2) and (4) are Log Employees. Models (1) and (4) with standard errors adjusted for 22 clusters in strata, models (2) and (5) with standard errors adjusted for 30,961 clusters in firm ID, and models (3) and (6) with standard errors adjusted for 61 clusters in NACE code (2-digit level). + p<0.10; * p<0.05.

Table A6. Estimated Treatment Effects vs no treatment

Year	Dep. Variable	Audit	SE	Letter	SE
2017	Log Payroll Tax	0.200	(0.117)	0.109*	(0.039)
2018	Log Payroll Tax	0.393*	(0.188)	0.127**	(0.041)
2019	Log Payroll Tax	0.343+	(0.174)	0.219**	(0.075)
2020	Log Payroll Tax	0.397*	(0.184)	0.150*	(0.069)
2017	Log Employees	-0.002	(0.022)	0.035*	(0.011)
2018	Log Employees	0.035	(0.031)	0.039**	(0.011)
2019	Log Employees	0.083*	(0.038)	0.055***	(0.013)
2020	Log Employees	0.085*	(0.039)	0.053***	(0.011)

Note: The coefficients are to be interpreted in log-points. Fixed effects on strata and year on Audit regressions. Fixed effects on year on Letter regressions. Standard errors are adjusted for 22 clusters in strata. N Audits = 22 961. N Letters = 28 987. + p<0.10; * p<0.05; ** p<0.01; *** p<0.001.

Table A7. Estimated ITT vs LATE Letter Effects

Year	Dep. Variable	ITT	SE	LATE	SE
2017	Log Payroll tax	0.109*	(0.039)	0.127**	(0.045)
2018	Log Payroll tax	0.127**	(0.041)	0.148**	(0.050)
2019	Log Payroll tax	0.219**	(0.075)	0.255**	(0.088)
2020	Log Payroll tax	0.150*	(0.069)	0.174*	(0.082)
2017	Log Employees	0.035*	(0.011)	0.040	(0.015)
2018	Log Employees	0.039**	(0.011)	0.046***	(0.014)
2019	Log Employees	0.055***	(0.013)	0.064***	(0.015)
2020	Log Employees	0.053***	(0.011)	0.062***	(0.013)

Note: The coefficients are to be interpreted in log-points. Fixed effects on year. Standard errors are adjusted for 22 clusters in strata. N = 28 987. + p<0.10; * p<0.05; ** p<0.01; *** p<0.001.

9.7 Heterogeneous Effects

Table A8. Estimated Audit and Letter (ATE) Effects (base year 2017) on firm size 25 percentile

Treatment	(1)	(2)	(3)	(4)	(5)	(6)
Audit*2018	-0.161	-0.161	0.124	-0.011	-0.011	-0.011
	(0.265)	(0.201)	(0.079)	(0.018)	(0.015)	(0.021)
Audit*2019	-0.058	-0.058	0.041	-0.002	-0.002	-0.002
	(0.364)	(0.235)	(0.101)	(0.024)	(0.020)	(0.030)
Audit*2020	0.178	0.178	0.099	0.009	0.009	0.009
	(0.344)	(0.242)	(0.108)	(0.025)	(0.024)	(0.029)
Letter*2018	-0.091	-0.091	0.018	-0.002	-0.002	-0.002
	(0.136)	(0.114)	(0.034)	(0.007)	(0.008)	(0.007)
Letter*2019	-0.010	-0.010	0.110*	0.001	0.001	0.001
	(0.157)	(0.135)	(0.049)	(0.009)	(0.012)	(0.009)
Letter*2020	-0.187	-0.187	0.040	-0.007	-0.007	-0.007
	(0.157)	(0.143)	(0.043)	(0.012)	(0.014)	(0.014)
Observations	30711	30711	123844	29439	29439	29439
R-squared	0.826	0.826	0.756	0.922	0.922	0.922

Note: Regressions with Fixed effects on strata and firm ID. The coefficients are to be interpreted in log-points. Standard errors in parenthesis. Columns (1) - (3) are estimated coefficients on Log Payroll Tax for firms with Revenue < NOK 735.003 and Columns (2) and (4) are estimated coefficients on Log Employees for firms with Revenue > NOK 9.200.000. Models (1) and (4) with standard errors adjusted for 22 clusters in strata, models (2) and (5) with standard errors adjusted for 30,961 clusters in firm ID, and models (3) and (6) with standard errors adjusted for 61 clusters in NACE code (2-digit level). + p<0.10; * p<0.05.

Table A9. Estimated Audit and Letter (ATE) Effects (base year 2017) on firm type Private Ltd

Treatment	(1)	(2)	(3)	(4)	(5)	(6)
Audit*2018	0.191+	0.191**	0.191*	0.042+	0.042***	0.042*
	(0.093)	(0.069)	(0.092)	(0.021)	(0.011)	(0.020)
Audit*2019	0.070	0.070	0.070	0.078**	0.078***	0.078**
	(0.094)	(0.104)	(0.105)	(0.024)	(0.019)	(0.025)
Audit*2020	0.077	0.077	0.077	0.045*	0.045+	0.045+
	(0.094)	(0.119)	(0.099)	(0.020)	(0.025)	(0.025)
Letter*2018	0.020	0.020	0.020	0.005	0.005	0.005
	(0.031)	(0.043)	(0.033)	(0.005)	(0.007)	(0.005)
Letter*2019	0.130+	0.130*	0.130*	0.022+	0.022+	0.022+
	(0.064)	(0.058)	(0.055)	(0.011)	(0.012)	(0.012)
Letter*2020	0.088	0.088	0.088	0.025+	0.025+	0.025+
	(0.070)	(0.066)	(0.058)	(0.013)	(0.014)	(0.013)
Observations	131460	131460	131460	131460	131460	131460
R-squared	0.654	0.654	0.654	0.771	0.771	0.771

Note: Regressions with Fixed effects on strata and firm ID. The coefficients are to be interpreted in log-points. Standard errors in parenthesis. Columns (1) - (3) are estimated coefficients on Log Payroll Tax and Columns (2) and (4) are estimated coefficients on Log Employees. Models (1) and (4) with standard errors adjusted for 22 clusters in strata, models (2) and (5) with standard errors adjusted for 30,961 clusters in firm ID, and models (3) and (6) with standard errors adjusted for 61 clusters in NACE code (2-digit level). + p<0.10; * p<0.05.

Table A10. Estimated Audit and Letter (ATE) Effects (base year 2017) on firm type Self-employed

Treatment	(1)	(2)	(3)	(4)	(5)	(6)
Audit*2018	-0.183	-0.183	-0.183	-0.072+	-0.072+	-0.072*
	(0.285)	(0.213)	(0.286)	(0.036)	(0.040)	(0.035)
Audit*2019	0.025	0.025	0.025	-0.103*	-0.103+	-0.103*
	(0.418)	(0.270)	(0.411)	(0.048)	(0.054)	(0.048)
Audit*2020	0.377	0.377	0.377	-0.058	-0.058	-0.058
	(0.344)	(0.281)	(0.338)	(0.052)	(0.060)	(0.048)
Letter*2018	0.008	0.008	0.008	0.010	0.010	0.010
	(0.128)	(0.120)	(0.114)	(0.024)	(0.020)	(0.019)
Letter*2019	-0.079	-0.079	-0.079	0.007	0.007	0.007
	(0.182)	(0.157)	(0.134)	(0.037)	(0.029)	(0.033)
Letter*2020	-0.273	-0.273	-0.273*	-0.018	-0.018	-0.018
	(0.214)	(0.172)	(0.116)	(0.030)	(0.033)	(0.026)
Observations	22085	22085	22085	22085	22085	22085
R-squared	0.661	0.661	0.661	0.798	0.798	0.798

Note: Regressions with Fixed effects on strata and firm ID. The coefficients are to be interpreted in log-points. Standard errors in parenthesis. Columns (1) - (3) are estimated coefficients on Log Payroll Tax and Columns (2) and (4) are estimated coefficients on Log Employees. Models (1) and (4) with standard errors adjusted for 22 clusters in strata, models (2) and (5) with standard errors adjusted for 30,961 clusters in firm ID, and models (3) and (6) with standard errors adjusted for 61 clusters in NACE code (2-digit level). + $p<0.10$; * $p<0.05$.

Table A11. Estimated Audit and Letter (ATE) Effects (base year 2017) on firm type Norwegian registered foreign company

Treatment	(1)	(2)	(3)	(4)	(5)	(6)
Audit*2018	0.009	0.009	0.009	-0.478	-0.478	-0.478
	(1.179)	(1.361)	(1.285)	(0.283)	(0.323)	(0.316)
Audit*2019	-1.474	-1.474	-1.474	-0.339	-0.339	-0.339
	(1.292)	(1.719)	(1.454)	(0.248)	(0.308)	(0.289)
Audit*2020	-1.457	-1.457	-1.457	-0.262	-0.262	-0.262
	(1.462)	(1.589)	(1.556)	(0.266)	(0.345)	(0.291)
Letter*2018	0.003	0.003	0.003	-0.051	-0.051	-0.051
	(0.588)	(0.580)	(0.600)	(0.055)	(0.062)	(0.054)
Letter*2019	1.345	1.345	1.345*	0.062	0.062	0.062
	(0.832)	(0.815)	(0.573)	(0.132)	(0.113)	(0.136)
Letter*2020	0.589	0.589	0.589	0.011	0.011	0.011
	(1.086)	(0.932)	(0.799)	(0.174)	(0.153)	(0.182)
Observations	1260	1260	1260	1260	1260	1260
R-squared	0.685	0.685	0.685	0.768	0.768	0.768

Note: Regressions with Fixed effects on strata and firm ID. The coefficients are to be interpreted in log-points. Standard errors in parenthesis. Columns (1) - (3) are estimated coefficients on Log Payroll Tax and Columns (2) and (4) are estimated coefficients on Log Employees. Models (1) and (4) with standard errors adjusted for 22 clusters in strata, models (2) and (5) with standard errors adjusted for 30,961 clusters in firm ID, and models (3) and (6) with standard errors adjusted for 61 clusters in NACE code (2-digit level). + p<0.10; * p<0.05.

10. References

Advani, A., Elming, W., & Shaw, J. (2018). The dynamic effects of tax audits. Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association,

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

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A Theoretical Analysis. *Journal of Public Economics*, *1*(3-4), 323-338.

Alm, J. (2019a). Tax compliance and administration. In *Handbook on taxation* (pp. 741-768). Routledge.

Alm, J. (2019b). What motivates tax compliance? *Journal of economic surveys*, *33*(2), 353-388.

Almunia, M., & Lopez-Rodriguez, D. (2018). Under the radar: The effects of monitoring firms on tax compliance. *American Economic Journal: Economic Policy*, *10*(1), 1-38.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, 47-57.

Baer, M. K., & Silvani, M. C. (1997). *Designing a tax administration reform strategy: Experiences and guidelines*. International Monetary Fund.

Beer, S., Kasper, M., Kirchler, E., & Erard, B. (2020). Do audits deter or provoke future tax noncompliance? Evidence on self-employed taxpayers. *CESifo Economic Studies*, *66*(3), 248-264.

Bergolo, M., Ceni, R., Cruces, G., Giaccobasso, M., & Perez-Truglia, R. (2023). Tax audits as scarecrows: Evidence from a large-scale field experiment. *American Economic Journal: Economic Policy*, *15*(1), 110-153.

Bjørneby, M., Alstadsæter, A., & Telle, K. (2021). Limits to third-party reporting: Evidence from a randomized field experiment in Norway. *Journal of Public Economics*, *203*, 104512.

Blumenthal, M., Christian, C., & Slemrod, J. (2001). *Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in minnesota*. https://EconPapers.repec.org/RePEc:feb:natura:00332

Boning, W. C., Guyton, J., Hodge, R., & Slemrod, J. (2020). Heard it through the grapevine: The direct and network effects of a tax enforcement field experiment on firms. *Journal of Public Economics*, *190*, 104261.

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

Brockmeyer, A., Smith, S., Hernandez, M., & Kettle, S. (2019). Casting a wider tax net: Experimental evidence from Costa Rica. *American Economic Journal: Economic Policy*, *11*(3), 55-87.

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.

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

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

Dhami, S., & Al-Nowaihi, A. (2007). Why do people pay taxes? Prospect theory versus expected utility theory. *Journal of Economic Behaviour & Organization*, *64*(1), 171-192.

Doerrenberg, P., & Schmitz, J. (2015). Tax compliance and information provision–A field experiment with small firms. *ZEW-Centre for European Economic Research Discussion Paper*(15-028).

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

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behaviouralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, *148*, 14-31.

Hebous, S., Jia, Z., Løyland, K., Thoresen, T. O., & Øvrum, A. (2020). Do audits improve future tax compliance in the absence of penalties? Evidence from random Audits in Norway.

Kausar, A., Shroff, N., & White, H. (2016). Real effects of the audit choice. *Journal of Accounting and Economics*, *62*(1), 157-181.

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

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. *79*(3), 651-692. https://doi.org/10.3982/ecta9113

Kotsadam, A., Løyland, K., Raaum, O., Torsvik, G., & Øvrum, A. (2021). 12th Norwegian German Seminar on Public Economics.

Liang, K.-Y., & Zeger, S. L. (1993). Regression analysis for correlated data. *Annual review of public health*, *14*(1), 43-68.

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

Løyland, K., Raaum, O., Torsvik, G., & Øvrum, A. (2019). Compliance effects of risk-based tax audits.

Meiselman, B. S. (2018). Ghostbusting in Detroit: Evidence on nonfilers from a controlled field experiment. *Journal of Public Economics*, *158*, 180-193.

Micci-Barreca, D., & Ramachandran, S. (2004). Improving tax administration with data mining. *White paper. Elite Analytics LLC*.

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

Murphy, K. (2019). *Moving towards a more effective model of regulatory enforcement in the Australian Taxation Office*. Centre for Tax System Integrity (CTSI), Research School of Social Sciences ….

[Record #207 is using a reference type undefined in this output style.]

Ortega, D., & Scartascini, C. (2020). Don't blame the messenger. The Delivery method of a message matters. *Journal of Economic Behaviour & Organization*, *170*, 286-300.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, *105*(8), 2539-2569.

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

Slemrod, J. (2019). Tax compliance and enforcement. *Journal of economic literature*, *57*(4), 904-954.

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human resources*, *50*(2), 301-316.

Tversky, A., & Kahneman, D. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-291.

Weiss, M. J., Lockwood, J., & McCaffrey, D. F. (2016). Estimating the standard error of the impact estimator in individually randomized trials with clustering. *Journal of Research on Educational Effectiveness*, *9*(3).