INSTITUTE OF RETAIL ECONOMICS

REPLICATING AND EXTENDING 'VOLUNTARY VS MANDATORY: THE ROLE OF AUDITING IN CONSTRAINING CORPORATE TAX AVOIDANCE IN SMALL PRIVATE FIRMS'

FREDRIK HARTWIG, ASIF M. HUQ & NIKLAS RUDHOLM

HFI WORKING PAPER No 41

Replicating and extending 'Voluntary vs mandatory: the role of auditing in constraining corporate tax avoidance in small private firms'

Fredrik Hartwig, Asif M Huq, Niklas Rudholm

Abstract: Many European countries have abolished mandatory audits for small firms to reduce their regulatory burden. In a recent article, Dong et al. (2023) reported that voluntarily audited firms experienced a 19% reduction in their total tax burden compared to mandatorily audited firms following an auditing reform in Sweden, suggesting a causal link between the revised auditing regulation and corporate tax avoidance. In this study, we first replicate Dong et al.'s (2023) results with reasonable accuracy. We then incorporate a series of placebo tests to examine whether their results are causal effects of the reform or spurious correlations. Placebo tests adjusting the timing of the reform, along with tests modifying the size of firms eligible for voluntary audits under the reform, reveal statistically significant reform effects where none should be expected. To further investigate, we conduct an independent analysis of the data, finding that voluntarily audited firms increased their tax payments due to the introduction of the reform, rather than reducing them.

Keywords: Tax evasion, regulatory reform, earnings management, placebo testing, conditional difference-in-difference analysis.

JEL-codes: M42, M48, D22

1. Introduction

Research is often used by regulators or policymakers to guide business or policy decisions that can affect firms and individuals. Therefore, it is crucial that these decisions are grounded in scientific evidence, which can only be achieved through rigorous verification and replication of pioneering empirical studies. Without corroboration, the findings of individual studies cannot provide the knowledge necessary to make confident policy prescriptions (Babin et al., 2021).

In recent years, empirical studies have analysed regulatory changes that make external audits voluntary for small firms. Due to concerns regarding the regulatory burden on smaller firms, most European countries have exempted these firms from mandatory audits. The argument is that the costs associated with auditing small firms are often disproportionality high relative to the limited benefits received (Keasey et al., 1988). Empirical evidence also suggests that exempting smaller firms from audits leads to higher employment growth among these firms (Huq et al., 2021). Other studies present evidence that private firms that opted for voluntary audits experienced lower interest rates on bank loans (Dedman et al., 2014; Huq et al., 2022; Kim et al., 2011) and were more likely to receive credit (Alduraywish, 2023) compared to those that did not undergo audits. Conversely, Koren et al. (2014) reported that firms with voluntarily audited financial statements had a higher cost of debt compared to unaudited firms. The relationship between audit status and accounting quality has also been examined. For example, Clatworthy and Peel (2013) studied a sample of UK private firms and found that those which underwent voluntary audits reported fewer accounting errors compared to firms which forwent audits. Similarly, Downing and Langli (2019) highlighted a decline in reporting quality among firms that chose to forgo audits – compared to firms that chose to continue to be audited – following Norway's statutory auditing deregulation for small firms in 2011. Additionally, Dedman and Kausar (2012) demonstrated that firms which retained a voluntary audit were associated with more conservative financial reporting than those which opted out of audits.

One of the most important issues for policymakers regarding the regulatory reforms that make auditing voluntary for smaller firms is their impact on corporate tax avoidance. In a recent study, Dong et al. (2023) used conditional difference-in-difference (DiD) analysis to compare voluntarily and mandatorily audited firms following a regulatory reform in Sweden in 2010 that allowed voluntary audits for micro-firms.[1] Their study found that voluntarily audited firms

[1] In conditional DiD, matching procedures are first used to find valid control groups (e.g., Heckman et al., 1998).

exhibited a significant reduction in their total corporate income tax burden compared to their mandatorily audited counterparts, leading the authors to conclude that 'the shift to voluntary audits facilitates a higher level of tax avoidance' (Dong et al., 2023, p. 725).

The purpose of our research is to conduct a replication study to corroborate and extend the findings of Dong et al. (2023). Specifically, we first conduct an independent reproduction replication study, as defined by Babin et al. (2021) of the main empirical analysis presented in Dong et al.'s (2023) study. The objective of a reproduction replication is to keep the conditions of the original analysis – including samples, procedures, and methods – as identical to the original article as possible to corroborate its findings (Babin et al., 2021). Ideally, this type of replication should yield results that match those of the original article.

Next, we conduct what Babin et al. (2021, p. 507) term a differentiated replication, incorporating extensions of Dong et al.'s (2023) study. In a differentiated replication, researchers deliberately modify methods, conditions, or research procedures to corroborate the results of a previous study. Beyond corroborating the original study, this type of replication also aims to investigate whether its results hold under alternative specifications.

Our study's reproduction replication results show that, with reasonable accuracy, we can replicate the main results of Dong et al. (2023). In their study, voluntarily audited firms, on average, reduced their total tax burden, $TOTAL\_TAX_{ft}$, by 0.5 percentage points, and also reduced their conforming tax avoidance, $CONFTAX_{ft}$, by 0.5 percentage points relative to their mandatorily audited counterparts.[2] Dong et al. (2023) also found no significant impact of the regulatory reform on the effective tax rate, ETR_{ft}. Our replication estimate of the impact of the reform on the total tax burden ($TOTAL\_TAX_{ft}$) is slightly smaller, with a reduction of 0.2 percentage points, while we observe a 0.1 percentage point reduction in the conforming tax avoidance measure ($CONFTAX_{ft}$) and a statistically significant 1.2 percentage point increase in the effective tax rate (ETR_{ft}) for voluntarily audited firms.

However, the results also show that placebo tests, conducted by manipulating reform dates both before and after the actual reform, also reveal statistically significant reform effects. These significant results could still be attributed to the reform for the placebo tests conducted after the reform if it had created a new trend, rather than a new steady-state level, for the outcome variables. Explaining the statistically significant pre-reform placebo test is more challenging

[2] Exact definitions of all outcome variables are provided in Section 3.1.2 below.

and would suggest that the trends of the outcome variables in the treatment and control groups were not parallel during the pre-reform period.

Following the placebo tests across time, we also performed a placebo test where the size thresholds of the regulatory reform were manipulated. These estimations only include data from the control group of the original estimations, meaning that there should be no statistically significant reform effects in this sample of firms. Nevertheless, these estimations show statistically significant placebo effect estimates as well.

Since our results indicate difficulties in verifying the validity of the parallel trend assumption for difference-in-difference (DiD) estimations using the methods suggested by Dong et al. (2023), we also employ an alternative method of analysis. This method is a step-by-step approach, beginning with an investigation of the parallel trend assumption on the original, pre-matching datasets. The analysis shows that for the effective tax rate, ETR_{ft}, the original data display parallel trends in the pre-reform period without any matching, whereas for $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$, trends are not parallel. Consequently, we apply coarsened exact matching (CEM; Blackwell et al., 2009; Iacus et al., 2011, 2012) to the outcome variables $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$ in the pre-reform period to ensure that the parallel trend assumption holds. After matching, we were able to verify that the pre-reform trends are parallel for all three outcome variables, though the necessary matching significantly reduced the number of firms and observations used in the analysis for $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$. Therefore, inferences for these variables should only be made within the sample.

The results from estimations using the datasets, where we can show that pre-reform trends are indeed parallel, show the following. First, it should be noted that in this analysis, tax payments – measured either as $TOTAL\_TAX_{ft}$, $CONFTAX_{ft}$ or ETR_{ft} – all increased in the voluntarily audited firms in the post-reform period compared to their tax payments in the pre-reform period and those of mandatorily audited firms over the entire period under study. Over the 4-year post-reform period, the average treatment effect is 0.002 for $TOTAL\_TAX_{ft}$, 0.001 for $CONFTAX_{ft}$, and 0.008 for ETR_{ft}, all statistically significant at the 1% level. The increase in total taxes paid amounts to approximately 3,700 SEK per firm per year on average when using the results for

the $TOTAL\_TAX_{ft}$ variable.[3] For ETR_{ft}, which has greater external validity, the results instead show an increase in taxes paid of approximately 2,100 SEK.

The remainder of the paper is organized as follows. Section 2 presents the regulatory reform under investigation and results from previous studies on the Swedish reform. Section 3 presents the reproduction replication study, beginning by presenting the data, sample construction, and research design and ending with a comparison of our empirical results to those of Dong et al. (2023). In Section 4, we perform an extended replication, where we estimate year-by-year treatment effects and then conduct similar placebo tests on our data, as conducted in the Dong et al. (2023) replication analysis. Finally, Section 5 summarises our results and then discusses insights from our research.

2. The Swedish voluntary audit reform

The EU Fourth Company Law Directive (78/660/EEC) grants EU member states the option to exempt small and medium-sized enterprises (SMEs) from mandatory audits (European Economic Community, 1978). Most countries within the EU also use this option to exempt SMEs from audits. Sweden, however, was a rare exception for a long time, requiring audits for all Swedish limited firms, even the smallest ones. The Swedish legislation can be traced back to 1895, when the Companies Act was re-written, mandating all limited firms to appoint independent auditors (Öhman & Wallerstedt, 2012).[4]

In 2006, a centre-right government was elected in Sweden. On 14 April 2010, the government submitted the bill 'A Voluntary Audit' (Prop. 2009/10:204) to the Swedish parliament, proposing that small firms be allowed to decide whether to appoint an auditor. The reform was justified on the grounds that the government aimed to reduce the regulatory burden on small firms, aligning with the European Commission's plan of reducing the administrative burden on SMEs by 20%.[5] The Swedish Parliament passed the bill on 21 June 2010 (SFS 2010:834), allowing firms below certain thresholds to be exempted from statutory audits from fiscal years beginning on or after 1 November 1 2010 (SFS 2010:834). The reform was expected to exempt

[3] Approximately 370 EURO. During the study period, the SEK/EURO exchange rate fluctuated between 8.89 SEK/EURO in 2014 and 10.65 SEK/EURO in 2009. We therefore consider 10 SEK/EURO a reasonable approximation for the years under study.

[4] Voluntary audits date back to the 1650s in Sweden. Official, though still voluntary, audits were also mentioned in the Companies Act of 1848.

[5] As per the EU Fourth Company Law Directive (78/660/EEC). Reports from the European Commission highlighted the importance of SMEs to the European economy, calling for a more business-friendly environment for SMEs, including micro firms, to enhance their competitiveness in the global economy.

approximately 72% of all Swedish limited firms from mandatory audits (Svanström & Sundgren, 2012).

While the Swedish Companies Act (*Aktiebolagslagen*, 2005:551) Chapter 9 §1 continues to formally stipulate that – as a starting point – all Swedish limited firms must have at least one auditor, privately owned, limited firms' Articles of Associations may, from 1 November 2010, specify an exemption from this requirement if least two of the following three conditions are met:[6]

i. The average number of employees for the previous 2 consecutive fiscal years amounts to no more than three.
ii. Reported total assets for the previous 2 consecutive fiscal years amounts to no more than 1.5 million SEK.
iii. Reported net sales for the previous 2 consecutive fiscal years amounts to no more than 3 million SEK.

These thresholds mean that only micro firms can opt out of mandatory audits in Sweden. By contrast, most other European countries (with the exceptions of Finland, Malta, and Cyprus) set significantly higher thresholds for mandatory audits (Table 1).

The voluntary audit reform has been studied in four Swedish reports and two academic journal articles. First, the impact of voluntary auditing on economic crime was studied in two reports by the Swedish Economic Crime Authority (2016; 2020). These reports present statistics that claim that firms opting out of auditing are overrepresented among those reported for economic crime. However, Daunfeldt and Rudholm (2021) reviewed the first report and found that, out of a total of five calculations carried out in the Swedish Economic Crime Authority's report, four showed no overrepresentation of opted-out firms among those reported for suspected economic crimes. In the final calculation, the results indicated that firms opting out of audits were overrepresented among those reported for crimes. However, this calculation relied on a categorization of firms created by the Swedish Economic Crime Authority, and the results would lose statistical significance if just four out of 99 firms were misclassified within this categorisation.

[6]This also applies to the parent company of a group (even if the threshold levels in the parent company are not exceeded), provided the group meets more than one of the specified conditions. All intercompany claims and transactions, however, must first be eliminated.

Table 1. Threshold values for mandatory auditing in European countries, as of April 2021 (Accountancy Europe, 2021).

Country	Threshold: Total assets (EUR)	Threshold: Net turnover (EUR)	Threshold: Number of employees
Austria	5.000.000	10.000.000	50
Belgium	4.500.000	9.000.000	50
Bulgaria	1.000.000	2.000.000	50
Croatia	2.000.000	4.000.000	25
Cyprus	0	0	0
Czech Republic	1.500.000	3.000.000	50
Denmark	6.000.000	12.000.000	50
Estonia	2.000.000	4.000.000	50
Finland	100.000	200.000	3
France	4.000.000	8.000.000	50
Germany	6.000.000	12.000.000	50
Greece	4.000.000	8.000.000	50
Hungary	Not applicable	833.333	50
Iceland	1.400.000	2.800.000	50
Ireland	6.000.000	12.000.000	50
Italy	4.000.000	4.000.000	20
Latvia	800.000	1.600.000	50
Lithuania	1.800.000	3.500.000	50
Luxembourg	4.400.000	8.800.000	50
Malta	46.600	93.000	2
Netherlands	6.000.000	12.000.000	50
Norway	2.500.000	625.000	10
Poland	2.500.000	5.000.000	50
Portugal	1.500.000	3.000.000	50
Romania	3.350.000	6.650.000	50
Slovakia	2.000.000	4.000.000	30
Slovenia	4.000.000	8.000.000	50
Spain	2.850.000	5.700.000	50
Sweden	150.000	300.000	3
Switzerland	18.203.000	36.405.000	250
Turkey	4.070.000	8.140.000	175
United Kingdom	6.541.000	13.082.000	50

Note: The threshold for total assets does not apply in Hungary, as its audit obligation is determined solely by whether the thresholds for net turnover and the number of employees are exceeded.

In the second report (Swedish Economic Crime Authority, 2020), the purpose was to investigate whether 47 Swedish firms, whose board members were suspected of economic crime related to firm management, had opted out of auditing more frequently than firms in general. Daunfeldt and Rudholm (2021) found the report challenging to evaluate, as the methods section was limited to a single paragraph covering less than half a page. Nonetheless, based on the figures presented in the report, Daunfeldt and Rudholm (2021) demonstrated that there were no statistically significant differences in audit usage between the 47 sampled firms and firms in general at the time the investigation began. It should also be noted that none of the suspected individuals in these firms were ultimately convicted.

The reform's impact on firm growth has been studied in a report by the Swedish National Audit Office (2017) and two published scientific articles (Huq et al. 2021; 2022). The Swedish National Audit Office (SNAO, 2017) reported that firms opting out of auditing experienced lower sales growth after the reform compared to those that voluntarily retained an auditor, concluding that the reform did not lead to increased firm growth. However, an issue with the analysis is the potential for selection bias, as firms expecting to grow may choose to retain their auditor, anticipating that they will soon exceed the mandatory auditing thresholds. Despite this limitation, the SNAO compares two groups of firms, both given the option to opt out of auditing, which creates obvious problems in the analysis if those voluntarily retaining auditing have, on average, greater growth ambitions than those opting out.

The notion that firms voluntarily choosing audits of their financial statements are more growth-oriented is likely based on studies from other countries (Dedman et al., 2014; Ojala et al., 2016) and supported by the two scientific articles on the Swedish reform (Huq et al., 2021; 2022). Huq et al. (2021) found that firms permitted to opt out of auditing increased employment growth by 0.59% due to the reform, hence despite the Swedish reform's limited scope compared to similar reforms in other countries, approximately 2,800 new jobs were created. A possible reason for faster growth among firms allowed to opt out of audits is their ability to choose strategies optimal to their individual needs. Firms that opt out of auditing benefit from improved cash flow due to reduced costs, while Kim et al. (2011), Dedman et al. (2014), and Huq et al. (2022) have all shown that firms voluntarily choosing audits exhibit lower costs of capital in their financial reports. Both mechanisms free up resources for firm growth.

3. Reproduction replication of Dong et al. (2023)

In this section, we conduct a reproduction replication study, as defined by Babin et al. (2021), of the main empirical analysis presented in Dong et al. (2023). To ensure the analysis is as comparable and easy to follow as possible, the same structure, headlines, and tables as in the empirical part of Dong et al. (2023) are used to the extent possible, with additional information supplied in the Online supplemental material to this article.

3.1 Data, sample construction, and research design

3.1.1. Data sources and sample construction

In Sweden, all limited liability firms are required to submit annual reports containing financial statement data to the Swedish Companies Registration Office (SCRO). As in Dong et al.

(2023), our database includes information from all registered limited liability firms in Sweden. In addition to the financial statement data from SCRO, the dataset also includes information regarding the firms' auditors and other general firm information from Statistics Sweden. These data have been compiled by the Swedish House of Finance National Research Data Center and licenced to the University of Gävle for research.

The construction of the main dataset is presented in Appendix A1 in the Online supplemental material and can be described as follows. Our starting sample consists of 1,283,752 firm-years for 242,595 unique firms for the years 2006 to 2014. Following Dong et al. (2023), we include firms that were active in all 9 years (2006–2014), resulting in 68,271 unique firms and 614,439 firm-years, of which 537,952 are audited firm-years compared to 472,716 firm-years in Dong et al. (2023).

Next, we only include firms that maintained either a stable mandatory or voluntary audit over the study period. Thus, our [Dong et al. (2023) in square brackets] sample before matching consists of 11,146 [18,312] unique firms subject to stable mandatory audits, comprising 100,314 [164,808] firm-years. Additionally, 18,698 [12,924] unique firms maintaining stable voluntary audits, comprising 168,282 [116,319] firm-years.

The matched sample is obtained by matching 1:1 with a 0.25 calliper on sales growth ($SALES\_GR_{ft}$), total assets ($SIZE_{ft}$), and leverage ($LEVERAGE_{ft}$) for 2010, and exact matching on 2-digit SNI codes. Our matched sample consists of 3,843 [2,407] unique firms in each group, comprising 69,174 [43,326] firm-years. In Table 2, Section 3.2.1., we report descriptives of all variables used in the study after matching.

3.1.2 Key variable construction

Dong et al. (2023) study corporate tax avoidance using three main outcome variables: total tax burden ($TOTAL\_TAX_{ft}$), conforming tax avoidance ($CONFTAX_{ft}$), and the cash-effective tax rate (ETR_{ft}). The first measure, $TOTAL\_TAX_{ft}$, is defined as total taxes paid divided by lagged total assets and is configured to measure comprehensive corporate tax avoidance. In the words of Dong et al. (2023, page 731), 'By design, this measure captures the outcomes of all types of tax-avoidance decisions made – conforming and non-conforming, legal and illegal'.

However, notably, this metric also captures changes related to firm size, which can be independent of tax planning behaviour. This is particularly serious if the treatment and control group firms exhibit different trends in asset growth during the period under study. A higher

average growth rate in total assets in voluntarily audited firms compared to mandatorily audited firms would then be mistaken for increased tax avoidance in voluntarily audited firms after the reform, even when no such behaviour exists. Such an outcome could be due to firms with growth ambitions choosing to be voluntarily audited to a greater extent than firms in general, either because they believe they will soon reach the limits for mandatory audits (Ojala et al., 2016) or because voluntarily audited firms have access to investment capital at a lower cost compared to firms that opt out of auditing altogether (Dedman et al., 2014; Huq et al., 2022; Kim et al., 2011).

Then, to capture conforming tax avoidance (i.e., tax avoidance that reduces tax liabilities by reducing both book and taxable income by the same amounts), Dong et al. (2023) use a measure developed by Badetscher et al. (2019). Like the measure of total tax burden previously discussed, $CONFTAX_{ft}$ is also based on the ratio of taxes paid to lagged total assets. However, according to Dong et al. (2023), using taxes paid divided by lagged total assets captures both non-conforming and conforming tax avoidance. To remove the impact of non-conforming tax avoidance, the total tax measure is regressed on total book-tax differences, an indicator variable equal to 1 for negative book-tax differences, an interaction between these variables, and variables measuring both the level and change in net operating loss carryforwards. The residual from the estimation of this regression constitutes the conforming tax avoidance ($CONFTAX_{ft}$) measure, as suggested by Badetscher et al. (2019). However, since this measure of conforming tax avoidance is also based on the ratio of taxes paid to total assets in the previous year, it suffers from the same limitations as the total tax burden.

Finally, Dong et al. (2023) use the cash-effective tax rate (ETR_{ft}), calculated as the ratio between reported tax payments and income after financial items. According to Badetscher et al. (2019), it is one of the most commonly used tax avoidance measures.[7] To measure how a reform affects tax avoidance, this measure is likely the least problematic of the given alternatives. The introduction of voluntary auditing could affect firm behaviour regarding either the size of the tax payments or the firms' reported income after financial items (i.e., the size of the tax base). The effective tax rate captures both these possible behavioural changes in the firms.

[7] Following Dong et al. (2023), all continuous variables (outcome and control) are winsorized at the 1% and 99% levels, except for ETR, which is winsorized at 0 and 1. See Dong et al. (2023, Appendix) for details.

It is worth noting that if voluntarily audited firms' total assets grow faster than their mandatorily audited counterparts, the analysis will indicate an increase in tax avoidance in the voluntarily audited firms over time if using either $TOTAL\_TAX_{ft}$ or $CONFTAX_{ft}$ in the analysis, even if the firms have not changed the size of either the tax payments or tax base throughout the period under study. However, for the effective tax rate, ETR_{ft}, firms must have changed their behaviour regarding the size of either the tax payments or tax base for a change in this tax avoidance measure to be observed. Consequently, if voluntarily audited firms grow faster in terms of assets compared to their mandatorily audited counterparts, without this directly affecting profits, estimations of the reform effect would be expected to yield negative estimates for $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$. Meanwhile, no effects would generally be identified when using ETR_{ft}, as this measure relates tax payments to profits rather than assets.

All these variables are measured on the firm (index *f*) and year (index *t*) levels, and by using all three measures, Dong et al. (2023) aim to provide a more complete analysis of the tax avoidance behaviour of firms and their managers compared to previous studies. When constructing the outcome variables, we follow Dong et al. (2023) in as much detail as possible, and a comparison of descriptive statistics shows that our data used in the estimations are quite similar (see Section 3.2.1 below).

3.1.3 Research design

In applied research, the DiD estimator is one of the most frequently used tools for evaluating the effects of interventions on relevant outcome variables (Abadie, 2005). The fundamental identification assumption in the DiD model presented in Equation (1) is that, in the absence of treatment, the outcome variable for firms in the treatment and control groups would have exhibited parallel trends. Following Dong et al. (2023), the following DiD model is estimated on the matched sample described above:

$$TAX_{ft} = \alpha + \beta_1 VOL\_AUDIT_f \times POST_t + \gamma' \mathbf{X} + \text{FIRM}_{\text{FE}} + \text{IND} \times \text{YEAR}_{\text{FE}} + \varepsilon_{ft} \qquad (1)$$

where TAX_{ft} represents either the total tax burden ($TOTAL\_TAX_{ft}$), conforming tax avoidance ($CONFTAX_{ft}$), or the cash-effective tax rate (ETR_{ft}). VOL_AUDIT_f is an indicator variable equal to 1 for firms belonging to the voluntarily audited group and 0 otherwise. $POST_t$ is an indicator variable equal to 1 for the years following the regulatory change (i.e., 2011 to 2014). We include the same set of firm-level control variables, **X**, as in Dong et al. (2023), along with

the same firm fixed effects (FIRM_FE) to capture time-invariant firm-level factors that could potentially bias the regulatory reform effect estimate, β_1. We also include the same industry-by-year fixed effects (IND × YEAR_FE) to capture potential differences in industry-specific time trends.

Our key variable of interest is the interaction between VOL_AUDIT_f and $POST_t$, as this will provide an estimate of the treatment effect. Specifically, how the different tax measures ($TOTAL\_TAX_{ft}$, $CONFTAX_{ft}$, or ETR_{ft}) of voluntarily audited firms compare with their own tax measures before the regulatory reform and with mandatorily audited firms throughout the study period, holding the levels of control variables and fixed effects constant. If the assumptions for DiD analysis are valid, a negative and statistically significant parameter estimate for β_1 would indicate that the tax payments of voluntarily audited firms have been reduced due to the regulatory reform.

3.2 Empirical results

3.2.1 Descriptive statistics

In Panel A of Table 2, we present descriptive statistics for our matched sample of 7,686 unique firms over the years 2006–2014, comprising 69,174 firm-years. Most importantly, the descriptives for the three measures of tax avoidance are close to those presented in Dong et al. (2023). With the means presented in Dong et al. (2023) in square brackets for comparison, we have the following: total taxes for the financial year scaled by total assets ($TOTAL\_TAX_{ft}$) 0.025 [0.027], conforming tax avoidance ($CONFTAX_{ft}$) 0.003 [0.001], and the effective tax rate (ETR_{ft}) 0.193 [0.200]. In Panels B (1) and B (2) of Table 2, we report the mean, standard deviation, and difference in the mean between the voluntary and mandatory audit groups for the pre- and post-reform years of the key variables reported by Dong et al. (2023). Finally, in Panel C of Table 2, we report the Pearson correlations of our matched sample for comparison to Dong et al. (2023).

Table 2. Descriptive statistics—Matched sample.

Panel A: Full sample for years 2006–2014 (N = 69,174)

	Mean	SD	Min	Median	Max
$TOTAL\_TAX_{ft}$	0.025	0.583	-6.691	0.011	152.800
$CONFTAX_{ft}$	0.003	0.023	-0.246	0.002	0.258
ETR_{ft}	0.193	0.211	0.000	0.199	1.000
$LARGE\_BOARD_{ft}$	0.292	0.455	0.000	0.000	1.000
Number of board members	2.532	1.262	1.000	2.000	31.000
$SIZE_{ft-1}$	8.028	0.999	3.638	8.021	10.631
Firm size, in SEK ´000	5,269	8,832	0.000	3,043	232,572
$PPERATIO_{ft-1}$	0.131	0.201	0.000	0.038	0.820
$CASH_{ft-1}$	0.232	0.230	0.000	0.162	0.999
$SALES\_GR_{ft-1}$	-0.022	0.620	-5.265	0.008	3.951
$LEVERAGE_{ft}$	0.113	0.203	0.000	0.000	0.859
$TRADE\_CREDIT_{ft}$	0.092	0.122	0.000	0.044	0.690
$MARGIN_{ft-1}$	-0.322	13.537	-2343.500	0.040	1.808
$RET\_EARNINGS_{ft-1}$	0.253	0.388	-3.694	0.229	0.988
NOL_{ft}	0.461	0.498	0.000	0.000	1.000
$LOSS_{ft}$	0.254	0.435	0.000	0.000	1.000
Firm Age, in years	21.716	12.577	3	20	146
$EXPOSED_{ft}$	0.008	0.087	0.000	0.000	1.000

Panel B (1): By group in the years 2006–2010 (N = 19,215 for each group)

	Voluntary audit firms 2006–2010		Mandatory audit firms 2006–2010		
	Mean	St. dev	Mean	St. dev	Difference
$TOTAL\_TAX_{ft}$	0.033	1.104	0.025	0.031	0.008
$CONFTAX_{ft}$	0.004	0.029	0.006	0.019	-0.002***
ETR_{ft}	0.191	0.221	0.227	0.215	-0.037***
$SIZE_{ft-1}$	7.678	1.114	8.307	0.682	-0.629***
$SALES\_GR_{ft-1}$	-0.021	0.837	0.020	0.221	-0.041***
$LEVERAGE_{ft}$	0.117	0.221	0.118	0.190	-0.001
$EXPOSED_{ft}$	0.003	0.057	0.012	0.111	-0.009***

Panel B (2): By group in the years 2011–2014 (N = 15,372 for each group)

	Voluntary audit firms 2011–2014		Mandatory audit firms 2011–2014		
	Mean	St. dev	Mean	St. dev	Difference
$TOTAL\_TAX_{ft}$	0.017	0.039	0.022	0.028	-0.005***
$CONFTAX_{ft}$	-0.000	0.024	0.002	0.017	-0.002***
ETR_{ft}	0.159	0.205	0.188	0.193	-0.028***
$SIZE_{ft-1}$	7.762	1.167	8.386	0.735	-0.624***
$SALES\_GR_{ft-1}$	-0.091	0.853	-0.010	0.270	-0.081***
$LEVERAGE_{ft}$	0.106	0.212	0.110	0.185	-0.004*
$EXPOSED_{ft}$	0.002	0.048	0.012	0.111	-0.010***

Panel C: Correlation matrix

Term	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) $TOTAL\_TAX_{ft}$	1															
(2) $CONFTAX_{ft}$	0.06*	1														
(3) ETR_{ft}	0.01*	0.23*	1													
(4) VOL_AUDIT_{f}	0.00	-0.04*	-0.08*	1												
(5) $POST_{i}$	-0.01*	-0.08*	-0.08*	0	1											
(6) $LARGE\_BOARD_{ft}$	0.01	0.01	0.01*	-0.18*	0	1										
(7) $SIZE_{ft-1}$	0.00	-0.04*	0.08*	-0.31*	0.04*	0.16*	1									
(8) $PPERATIO_{ft-1}$	0.00	-0.04*	-0.09*	-0.03*	-0.06*	-0.05*	-0.07*	1								
(9) $CASH_{ft-1}$	0.01*	0.16*	0.14*	0.1*	0.04*	-0.01	-0.09*	-0.22*	1							
(10) $SALES\_GR_{ft-1}$	0.00	0.06*	0.02*	-0.05*	-0.04*	0.02*	0.02*	0.04*	-0.05*	1						
(11) $LEVERAGE_{ft}$	-0.02*	-0.2*	-0.15*	0.00	-0.02*	0.02*	0.13*	0.17*	-0.38*	0.02*	1					
(12) $TRADE\_CREDIT_{ft}$	-0.01*	-0.01*	-0.04*	-0.3*	-0.04*	0.06*	-0.2*	-0.04*	-0.17*	0.05*	-0.04*	1				
(13) $MARGIN_{ft-1}$	0.00	0.00	0.02*	-0.03*	0.00	-0.01*	-0.02*	0.00	0.00	-0.02*	0.01*	0.01*	1			
(14) $RET\_EARNINGS_{ft-1}$	-0.04*	0.16*	0.11*	0.03*	0.05*	-0.04*	0.33*	-0.17*	0.2*	-0.04*	-0.32*	-0.29*	-0.02*	1		
(15) NOL_{ft}	-0.01*	-0.02*	-0.3*	0.09*	0.03*	-0.02*	-0.2*	0.11*	-0.22*	0.02*	0.15*	0.12*	-0.03*	-0.14*	1	
(16) $LOSS_{ft}$	-0.01*	0.00	-0.37*	0.16*	0.04*	-0.03*	-0.09*	0.00	-0.02*	-0.17*	-0.04*	-0.01*	-0.05*	0.09*	0.24*	1
(17) $EXPOSED_{ft}$	0.00	0.00	0.02*	-0.06*	0.00	0.04*	0.06*	-0.02*	0.01*	0.01*	-0.02*	0.02*	0.00	0.00	-0.01	-0.01*

Notes: This table presents summary statistics of the matched sample. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.2.2 The parallel trends assumption

The fundamental identification assumption in the DiD model presented in Equation (1) is that, in the absence of treatment, the outcome variables for firms in both the treatment and control groups would have exhibited parallel trends in the post-reform period.[8] Of course, the development of TOTAL_TAX, CONFTAX, and ETR in the absence of Sweden's voluntary audit reform cannot be observed empirically in the years following the reform. However, we can observe the trends in these variables for both the treatment and control group firms in the years leading up to the reform. If the trends are parallel, this serves as an indication of the validity of the identification assumption.

Below, we follow Dong et al. (2023) and tabulate the 1-year differences in the outcome variables from 2007 to 2010, as well as the differences in this statistic between the treatment and control group firms. In Dong et al. (2023), the only statistically significant (at the 10% level) difference between the groups is for CONFTAX 2009. In our analysis, we find statistically significant differences for TOTAL_TAX 2007 and CONFTAX 2009. However, the differences are small, and if the determination of parallel trends had been judged solely on this table, the trends may have been deemed parallel.

It is important to note, however, that in Figure 1, trends do not appear to be parallel but instead seem to diverge both before and after the reform's implementation. This is also true for the pre-reform period when investigating Figure 1 in Dong et al. (2023, p. 737). These graphs then put the parallel trend assumption into question. Therefore, we will investigate this potential issue further by conducting placebo testing.

[8] In Dong et al. (2023, page 736), the identification assumption is misrepresented as being one of parallel trends before (rather than in the absence of) treatment.

Table 3. Voluntary vs. mandatory audit and corporate tax avoidance: Parallel trends assumption for the dependent variables.

	Mandatory Audit Firms	Voluntary Audit Firms	Difference	P-value	N	Year
Δ TOTAL_TAX$_{ft}$	0.001	-0.040	0.041	0.306	7686	2010
Δ TOTAL_TAX$_{ft}$	-0.002	0.037	-0.039	0.326	7686	2009
Δ TOTAL_TAX$_{ft}$	-0.004	-0.003	-0.001	0.248	7686	2008
Δ TOTAL_TAX$_{ft}$	0.001	-0.002	0.003**	0.031	7686	2007
Δ CONFTAX$_{ft}$	0.000	-0.000	0.001	0.273	7686	2010
Δ CONFTAX$_{ft}$	-0.001	-0.002	0.001*	0.073	7686	2009
Δ CONFTAX$_{ft}$	-0.002	-0.001	-0.001	0.109	7686	2008
Δ CONFTAX$_{ft}$	0.001	0.000	0.001	0.260	7686	2007
Δ ETR$_{ft}$	-0.001	-0.006	0.005	0.327	7686	2010
Δ ETR$_{ft}$	-0.018	-0.013	-0.005	0.420	7686	2009
Δ ETR$_{ft}$	-0.013	-0.016	0.003	0.624	7686	2008
Δ ETR$_{ft}$	0.001	0.003	-0.001	0.832	7686	2007

Notes: This table presents results from testing the parallel trends assumption for the differences-in-differences tests. * p < 0.1, ** p < 0.05, *** p < 0.01.

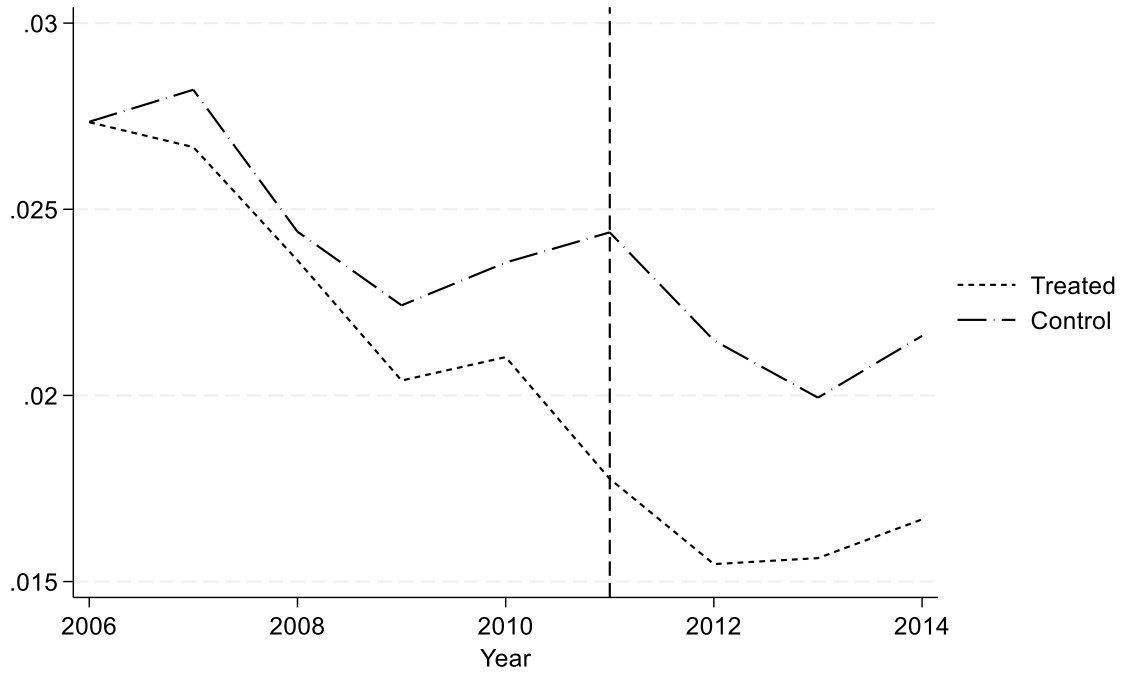


Figure 1. Tax scaled by assets ($TOTAL\_TAX_{ft}$) in voluntarily (treated) and mandatorily (control) audited firms. Dong et al. (2023) replication.

3.2.3 Effects of the voluntary audit reform on tax outcomes

As in Dong et al. (2023), we match the treatment and control group firms on sales growth ($SALES\_GR_{ft}$), total assets ($SIZE_{ft}$), and leverage ($LEVERAGE_{ft}$) in the year 2010. We also include firm and industry-year fixed effects to control for unobservable factors that could affect tax avoidance behaviour, and use the same set of control variables in the regression analysis. Although there are some differences in the descriptive statistics and the trends of the treatment and control groups, as presented in the graphs, the similarities are still substantial enough, meaning that our estimated treatment effects should closely resemble those presented in Dong et al. (2023).

The results from our estimations of Equation (1) are presented in Table 4 below, and they directly correspond to those presented in Table 4 of Dong et al. (2023). Our key variable of interest is the interaction between VOL_AUDIT_f and $POST_t$, as this will provide an estimate of how the different tax measures ($TOTAL\_TAX_{ft}$, $CONFTAX_{ft}$, or ETR_{ft}) of voluntarily audited firms compare with their own tax measures before the regulatory reform, and with those of mandatorily audited firms throughout the study period, holding the levels of the control variable and the fixed effects constant. If the identification assumption for DiD analysis of parallel trends in the absence of treatment is valid, a negative and statistically significant parameter estimate related to the interaction between VOL_AUDIT_f and $POST_t$ indicates that the tax payments of voluntarily audited firms – as measured by $TOTAL\_TAX_{ft}$, $CONFTAX_{ft}$, or ETR_{ft} – have been reduced due to the regulatory reform.

The results from Dong et al. (2023) show that voluntarily audited firms, on average, reduced their total tax burden (as measured by $TOTAL\_TAX_{ft}$) by 0.5 percentage points, which represents a 19% (0.005/0.027 = 0.19) decrease in tax payments, or approximately 15,000 SEK, relative to the mandatorily audited firms. Regarding the conforming tax avoidance measure, $CONFTAX_{ft}$, they report a reduction of 0.5 percentage points, while no significant impact of the regulatory reform on the effective tax rate, ETR_{ft}, was observed.

In our sample, the average tax payment in the years leading up to the reform (i.e., 2006–2010) is 95,455 SEK. However, voluntarily audited firms are smaller, hence they have lower profits and consequently pay less taxes. The average tax payments for the voluntarily audited firms equal 62,215 SEK, while the mandatorily audited firms pay, on average, 128,695 SEK per year in the pre-reform period. A 19% reduction in tax payments would imply a reduction in taxes paid of approximately 11,800 SEK (62,215 × 0.19 = 11,820) for the voluntarily audited firms. However, as shown in Table 4, our replication estimate of the impact of the reform on the total

tax burden is smaller than that reported by Dong et al. (2023), with a reduction of 0.2 percentage points, or 6% (0.002/0.033 = 0.06). This would correspond to a reduction in taxes paid of approximately 3,700 SEK (62,215 × 0.06 = 3,733). Regarding the conforming tax avoidance measure, $CONFTAX_{ft}$, we find a reduction of 0.1 percentage points, while also finding an increase in the effective tax rate, ETR_{ft}, for the voluntarily audited firms equal to 1.2 percentage points. It is worth noting that negative reform effect estimates for $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$, but not ETR_{ft}, would be expected if the voluntarily audited firms had a higher average growth rate in total assets compared to mandatorily audited firms (Section 3.1.2.).

Table 4. Replication of Dong et al. (2023), Table 4.

	Replication of Dong et al. (2023)		
	TOTAL TAX	CONFTAX	ETR
$VOL\_AUDIT_f \times POST_t$	-0.002***	-0.001***	0.012***
	(-5.27)	(-2.74)	(3.89)
$LARGE\_BOARD_{ft}$	0.000	-0.001	0.000
	(0.70)	(-1.45)	(0.07)
$SIZE_{ft-1}$	-0.014***	0.001	0.033***
	(-19.56)	(1.03)	(13.81)
$PPERATIO_{ft-1}$	0.000	0.001	0.025**
	(0.02)	(0.59)	(2.17)
$CASH_{ft-1}$	-0.002	0.001	0.020***
	(-1.19)	(0.93)	(3.01)
$SALES\_GR_{ft-1}$	0.002***	0.001***	-0.003***
	(9.22)	(5.47)	(-2.89)
$LEVERAGE_{ft-1}$	-0.024***	-0.015***	-0.095***
	(-15.31)	(-10.70)	(-10.62)
$TRADE\_CREDIT_{ft}$	-0.028***	-0.009***	-0.015
	(-11.11)	(-4.29)	(-1.18)
$MARGIN_{ft}$	0.000	0.000	0.000
	(-0.86)	(-1.12)	(1.46)
$RET\_EARNINGS_{ft-1}$	-0.007***	0.003***	-0.003
	(-5.71)	(2.67)	(-0.79)
NOL_{ft}	-0.004***	0.007***	-0.020***
	(-12.85)	(28.74)	(-8.83)
$LOSS_{ft}$	-0.016***	0.005***	-0.148***
	(-51.58)	(18.84)	(-61.06)
AGE_{ft}	-0.004**	-0.004***	0.031***
	(-2.05)	(-2.78)	(3.16)
Firm FE	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes
Observations	69,174	69,174	69,174
R^2	0.11	0.09	0.17

Note: t-values are reported in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01.

3.2.4 Placebo testing of our results

Placebo tests are designed to measure whether treatment effects on the dependent variables of interest – specifically, $TOTAL\_TAX_{ft}$, $CONFTAX_{ft}$, or ETR_{ft} – can be generated also by an invented or false treatment. Below, we perform placebo testing by manipulating and creating false reforms in terms of both the timing and size of reform thresholds. If the estimated effects

are truly caused by the reform, these placebo tests should result in no statistically significant reform effect estimates. However, if the results are instead driven by factors such as diverging rather than parallel trends in the outcome variables, statistically significant reform effects are also expected in the placebo estimations.

First, we perform two placebo tests across time, restricting the period of analysis to 5 years, and estimating the reform effect as if it had been introduced in 2008 (placebo before the true treatment) or 2014 (placebo after the true treatment).[9] In the estimations, we include all firms that have stable treatment or control status for the 5 (instead of 9) years under study. Otherwise, we identically edit the data as in the original replication estimations above.

If trends are parallel and the reform created a one-time shift in the outcome variables at its introduction, the pre- and post-reform placebo tests for 2008 and 2014 should yield statistically insignificant results for the treatment effect. However, if the trends in the outcome variables diverge over the period under study, placebo tests for 2008 and 2014 will show significant reform effects at points in time when no reforms were implemented.

The results in Table 5 show that the treatment effect estimate for the $TOTAL\_TAX_{ft}$ variable for the fictitious 2008 reform is negative, larger than the true reform year effect estimate (-0.002 in our replication), and statistically significant at the 1% level. The 2008 placebo effect estimate for the $CONFTAX_{ft}$ variable is of equal size to the true reform estimate and statistically significant at the 5% level, while the estimate for the ETR_{ft} variable is statistically insignificant at conventional levels. Thus, the placebo tests for $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$ in the pre-reform period yield results similar to those in the true treatment effect regressions, calling into question the findings of a 2010 causal effect of the reform.

Now, if there is an adjustment period by the firms in the years following the true reform, a DiD model could capture this adjustment and yield a statistically significant result in the post-reform (2014) placebo test. Thus, the post-reform period placebo test is weaker compared to the pre-reform period placebo test. Still, it is worth noting that for both the $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$ variables, the sizes of the placebo test reform effects are similar and statistically significant at the same level as the true reform effect estimations, thereby similarly calling into question the original results.

[9] For comparison, we also analyse the true 2010 reform using data from 2008 to 2012 (i.e., a 5-year window). In these regressions, we find a reform effect for TOTAL_TAX of -0.002***(-5.24)), for CONFTAX of -0.001***(-2.87), and for ETR of 0.006**(2.52).

If voluntarily audited firms have the same growth rate in taxes paid but a higher growth rate in total assets compared to mandatorily audited firms, the trends in the outcome variables will diverge, resulting in significant placebo effects identified in evaluations both before and after the true reform. In the data used for the replication, we find that the average growth rate in total assets over the period under study is nearly four times higher in the group of voluntarily audited firms (13% yearly growth) compared to those subject to mandatory auditing (3.6% yearly growth). The growth rate in taxes paid is also higher in the voluntarily audited firms, though only approximately 1.2 times higher than in the mandatorily audited firms, contributing to the divergent trends observed in the analysis.

The high growth rates for voluntarily and mandatorily audited firms in our sample are due to two selection mechanisms. First, voluntarily audited firms will grow faster than firms in general if firms with growth ambitions largely choose to be voluntarily audited. This may occur either because they expect to exceed the threshold for mandatory auditing soon (Ojala et al., 2016) or because voluntarily audited firms have access to investment capital at a lower cost (Dedman et al., 2014; Huq et al., 2022; Kim et al., 2011). Second, the sample consists entirely of firms that survived the entire 5- or 9-year periods under study. For small firms, quickly reaching minimum efficient scales of operation is crucial for survival (Audretch, 1995; Audretch et al., 2004; Daunfeldt et al., 2013). Many industries exhibit economies of scale, meaning that small firms are typically the first to exit the market during times of crisis or recession (Audretch, 1995; Audretch et al., 2004; Daunfeldt et al., 2013). Therefore, the small firms with the best chance of long-term survival are those that grow quickly and thus reach the necessary size.

If the faster growth of smaller surviving firms is responsible for the observed results, we can conduct another placebo test that is expected to yield results similar to those of the real reform. In this placebo test, we analyse a dataset consisting solely of mandatorily audited firms, where we divide the data into smaller firms just above the mandatory auditing threshold (treatment group) and slightly larger firms as the control group. If smaller firms generally have a higher growth rate in total assets than larger firms, we expect similar divergent trends in these data and therefore significant reform effects here as well.

In these placebo estimations, all firms that are in the treatment group (i.e., those below the voluntarily auditing thresholds set) in the original replication estimations are excluded. The variables determining whether a firm is in the placebo treatment group in the remaining data are based on exceeding the limits by 1 year, instead of the 2 years stipulated in the original reform, to avoid losing an excessive number of observations when using lagged variables in

the calculations. The following placebo conditions are then set for a firm to be classified as 'treated':

i. The average number of employees for the previous consecutive fiscal year amounts to more than three but less than six.
ii. Reported total assets for the previous consecutive fiscal year amounts to more than 1.5 but less than 3.0 million SEK.
iii. Reported net sales for the previous consecutive fiscal year amounts to more than 3.0 but less than 6.0 million SEK.

For the control group, we use all firms in the dataset above these thresholds with up to 10 employees. The results from these placebo test model estimations are presented in Table 6. The results show a placebo treatment effect that is larger than the true reform effect estimate for the $TOTAL\_TAX_{ft}$ variable, while the estimate for the $CONFTAX_{ft}$ variable is of similar size, and both estimates are statistically significant at the 1% level.

But this should be impossible, as all firms in this sample are required to mandatorily audit their financial reports and have done so for the entire period under study. Together with the time placebo results presented above, these findings indicate that the results presented in Dong et al. (2023) and Section 3.2.3 of this study are not due to the reform, but rather reflect differences in the trends of the outcome variables between firms in the treatment and control groups.

Table 5. Placebo tests manipulating the timing of the reform: 2008 and 2014.

	Placebo test [2008]			Placebo test [2014]		
	TOTAL_TAX	CONFTAX	ETR	TOTAL_TAX	CONFTAX	ETR
$VOL\_AUDIT_f \times POST_t$	−0.004***	−0.001*	0.005**	-0.002***	-0.001***	0.002
	(-9.68)	(-1.77)	(2.01)	(-6.46)	(-5.00)	(1.16)
$LARGE\_BOARD_{ft}$	0.001	0.000	-0.002	-0.000***	-0.001	0.004
	(0.62)	(0.10)	(-0.44)	(-0.44)	(-1.37)	(0.90)
$SIZE_{ft-1}$	-0.017***	0.004***	0.037***	-0.016***	0.004***	0.039***
	(-21.78)	(4.98)	(16.33)	(-22.82)	(5.10)	(18.30)
$PPERATIO_{ft-1}$	0.003	0.002	0.031***	-0.002	0.002	0.032***
	(1.45)	(1.06)	(2.82)	(-0.84)	(1.32)	(3.17)
$CASH_{ft-1}$	-0.008***	-0.000	0.029***	-0.003**	0.003***	0.030***
	(-6.03)	(-0.01)	(4.58)	(-2.36)	(3.19)	(5.81)
$SALES\_GR_{ft-1}$	0.002***	0.001***	-0.003***	0.001***	0.001***	-0.003***
	(7.89)	(4.70)	(-3.69)	(7.71)	(4.24)	(-2.81)
$LEVERAGE_{ft-1}$	-0.022***	-0.012***	-0.084***	-0.023***	-0.008***	-0.068***
	(-14.42)	(-7.68)	(-9.56)	(-14.24)	(-5.35)	(-8.38)
$TRADE\_CREDIT_{ft}$	-0.022***	-0.008***	-0.020	-0.031***	-0.003	0.005
	(-9.67)	(-3.98)	(-1.63)	(-12.97)	(-1.52)	(0.54)
$MARGIN_{ft}$	0.000	-0.000	0.000	0.000	-0.000	0.000
	(-0.62)	(-0.96)	(1.12)	(-0.84)	(-0.69)	(0.92)
$RET\_EARNINGS_{ft-1}$	-0.002**	0.005***	-0.007***	-0.000	0.005***	-0.006**
	(-2.35)	(3.99)	(-2.92)	(-0.17)	(4.75)	(-2.36)
NOL_{ft}	-0.003***	0.008***	0.001	-0.003***	0.008***	-0.001
	(-9.77)	(31.60)	(0.28)	(-10.52)	(37.62)	(-0.32)
$LOSS_{ft}$	-0.017***	0.005***	-0.154***	-0.015***	0.007***	-0.139***
	(-54.27)	(18.61)	(-65.48)	(-56.13)	(29.90)	(-70.58)
AGE_{ft}	-0.001	-0.002	0.059***	0.003*	-0.002*	0.001
	(-0.54)	(-1.27)	(4.84)	(1.76)	(-1.66)	(0.09)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	81,940	81,940	81,940	98,130	98,130	98,130
R^2	0.07	0.07	0.14	0.07	0.03	0.13

Note: t-values are reported in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01. Placebo estimations are performed on 5-year intervals around the placebo reform years of 2008 and 2014.

Table 6: Placebo test manipulating the size of the reform thresholds.

	Placebo test size		
	TOTAL TAX	CONFTAX	ETR
$VOL\_AUDIT_f \times POST_t$	-0.004***	-0.002***	0.012***
	(-8.18)	(-4.94)	(2.82)
$LARGE\_BOARD_{ft}$	0.001	0.001*	0.004
	(1.46)	(1.81)	(0.71)
$SIZE_{ft-1}$	-0.015***	-0.007***	0.038***
	(-16.89)	(-11.75)	(6.43)
$PPERATIO_{ft-1}$	-0.005**	-0.001	0.009
	(-2.04)	(-0.83)	(0.40)
$CASH_{ft-1}$	0.000	0.002*	0.031***
	(0.18)	(1.61)	(2.68)
$SALES\_GR_{ft-1}$	0.007***	0.004***	-0.003
	(8.63)	(8.60)	(-0.62)
$LEVERAGE_{ft-1}$	-0.023***	-0.013***	-0.099***
	(-13.58)	(-8.36)	(-6.68)
$TRADE\_CREDIT_{ft}$	-0.024***	-0.012***	-0.019
	(-10.08)	(-7.16)	(-0.92)
$MARGIN_{ft}$	0.019***	0.015***	0.040***
	(5.46)	(7.75)	(2.91)
$RET\_EARNINGS_{ft-1}$	-0.014***	-0.009***	-0.019*
	(-5.01)	(-4.66)	(-1.82)
NOL_{ft}	-0.003***	0.007***	-0.020***
	(-8.04)	(29.65)	(-5.62)
$LOSS_{ft}$	-0.012***	0.006***	-0.168***
	(-38.91)	(24.83)	(-45.67)
AGE_{ft}	-0.002	-0.001	0.043***
	(-1.07)	(-0.81)	(2.70)
Firm FE	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes
Observations	29,691	29,691	29,691
R^2	0.19	0.15	0.17

Notes: t-values are reported in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01. Placebo estimations are run on the control group used in the Dong et al. (2023) replication. The reform parameters are based on exceeding the limits by 1 year, instead of 2 years as stipulated in the original reform, to avoid losing an excessive number of observations due to lagged variables. The size thresholds for the treated group are as follows: SEK 3m < Sales <= SEK 6m, SEK 1.5m < total assets <= SEK 3m, and 3 < employees <= 6.

4. A differentiated replication with extensions

In this section, we perform what Babin et al. (2021) define as a differentiated replication, with extensions of the Dong et al. (2023) study. This type of replication deliberately modifies methods, conditions, or research procedures to corroborate the results from a previous study. The aim is not only to corroborate the original results but also to investigate whether they hold under varying conditions and methodologies.

4.1 Estimating year-by-year treatment effects using DiD

Since the placebo testing above indicates that the identification assumption of parallel trends is not valid after applying the matching approach used by Dong et al. (2023), an alternative method is required for conducting (conditional) DiD estimations of the reform's effect. To ensure both internal and external validity, the chosen method should satisfy the parallel trends identification assumption with minimal data loss (i.e., the data should as closely as possible represent the population of firms under investigation).

Thus, we propose the following method. We start with a dataset in which we have completed the basic data editing necessary, including winsorizing the variables, but before any type of matching. This dataset contains 268,596 observations, which is the sum of firm years for all firms with stable voluntary or mandatory audits (168,282 and 100,314 observations, respectively), as presented in Table A1.2 in Appendix A1 of the Online supplemental material. On these data, we then implement a more rigorous test of the parallel trend assumption than descriptive graphs, following Arcidiacono et al. (2020), modifying our estimating equation to include year-by-year treatment effect parameters. More formally, using the same control variables and fixed effects as in previous sections, we estimate the following year-by-year DiD model:

$$TAX_{ft} = \alpha + \sum_{t=2007}^{2014} \beta_t \left(VOL_{AUDIT_f} \times YEAR_t\right) + \gamma' \mathbf{X} + \mathrm{FIRM_{FE}}$$

$$+ \mathrm{IND} \times \mathrm{YEAR_{FE}} + \varepsilon_{ft} \qquad (2)$$

If the parallel trend assumption holds, the treatment effect estimates for the years leading up to the reform will be statistically insignificant and close to 0.[10] This is tested both by individual t-tests of the treatment effect parameter estimates for the years leading up to the reform and an F-test investigating whether the treatment effect parameter estimates for the years leading up to the reform are jointly significantly different from 0 at the 5% level. If they are, the parallel

[10] Estimating treatment effects for the years leading up to the reform serves as a placebo test, allowing us to investigate whether a treatment effect exists for years where none should occur.

trend assumption is questionable, and no causal interpretation of the post-reform treatment effect estimates can then be made.

The results from preliminary estimation of Equation (2) are presented in Appendix A2, Tables A2.1 (using data before any matching) and A2.2 (after coarsened exact matching, CEM; Blackwell et al., 2009; Iacus et al., 2011; 2012, on the different outcome variables using Sturges's rule to find the number of intervals, resulting in approximately 20 intervals) in the Online supplemental material, where the control variables are included in the estimations but excluded from the tables to save space. As shown in these tables, there are statistically significant treatment effects for the outcome variables $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$ in the years leading up to the reform. F-tests of the pre-reform year treatment effect parameters for these outcome variables are also statistically significant at the 5% level, invalidating the use of DiD estimations on these data. For ETR_{ft}, however, the estimations on the full dataset (Table A2.1), before any matching, show that the parallel trend assumption holds. Accordingly, no further action is taken on these data. The trends of ETR_{ft} for voluntarily (treated) and mandatorily (control) audited firms without matching are presented in Figure A2.1 in the Online supplemental material.

For the outcome variables $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$, we reduced the interval sizes permitted in the CEM matching by increasing the number of intervals to 50. We then repeated the analysis. Using year-by-year DiD estimations of the treatment effects on matched data with 50 intervals, we confirm (close to) parallel trends in the pre-reform years. This is also observed in a graph of the trends for $TOTAL\_TAX_{ft}$ after matching, which is presented in Figure 2.[11]

Although using such narrow intervals makes the treated and control group firms similar in terms of the development of the outcome variables in the years leading up to the reform, it also reduces the number of observations used in the estimations from 268,596 to 41,328 for $TOTAL\_TAX_{ft}$ and 66,762 for $CONFTAX_{ft}$. Accordingly, the DiD analysis on matched data is likely to perform well for internal validity, though at the expense of external validity. Therefore, any out-of-sample inference for the variables $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$ should be made with caution, while inferences using the ETR_{ft} variable results will be more representative of the population of firms under study.

[11] The corresponding figures for the $CONFTAX_{ft}$ and ETR_{ft} variables are provided in Appendix A3 of the Online supplemental material.

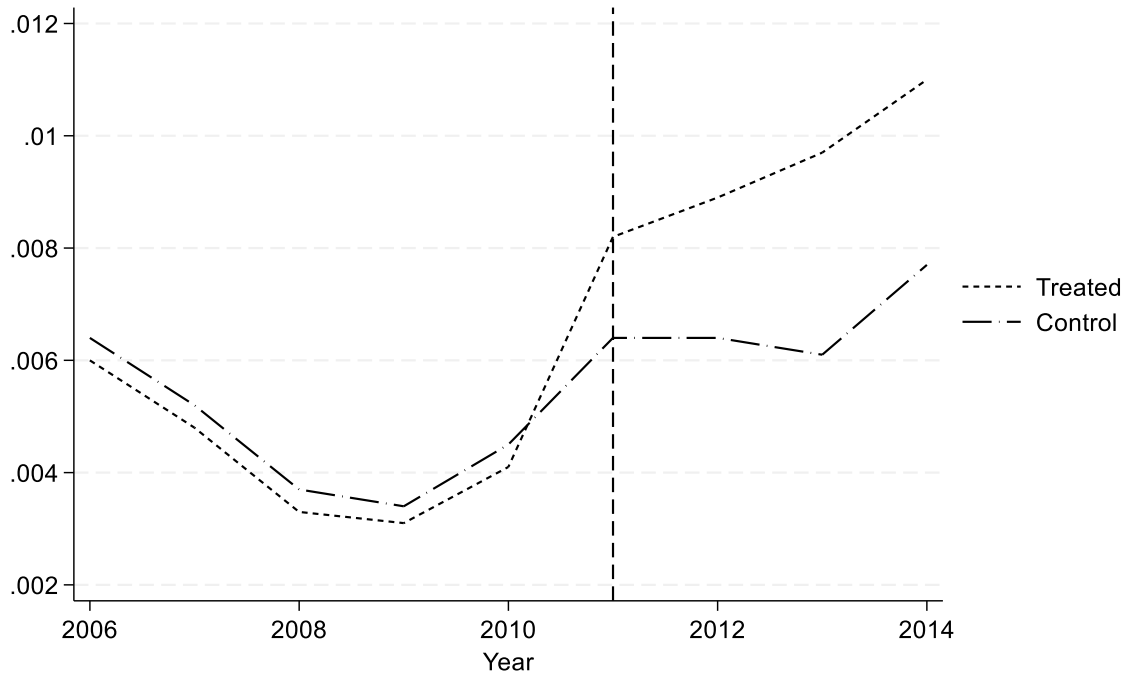


Figure 2. Tax scaled by assets ($TOTAL\_TAX_{ft}$) in voluntary (treated) and mandatorily (control) audited firms. CEM matching on $TOTAL\_TAX_{ft}$ for the years 2006 to 2010 using 50 intervals.

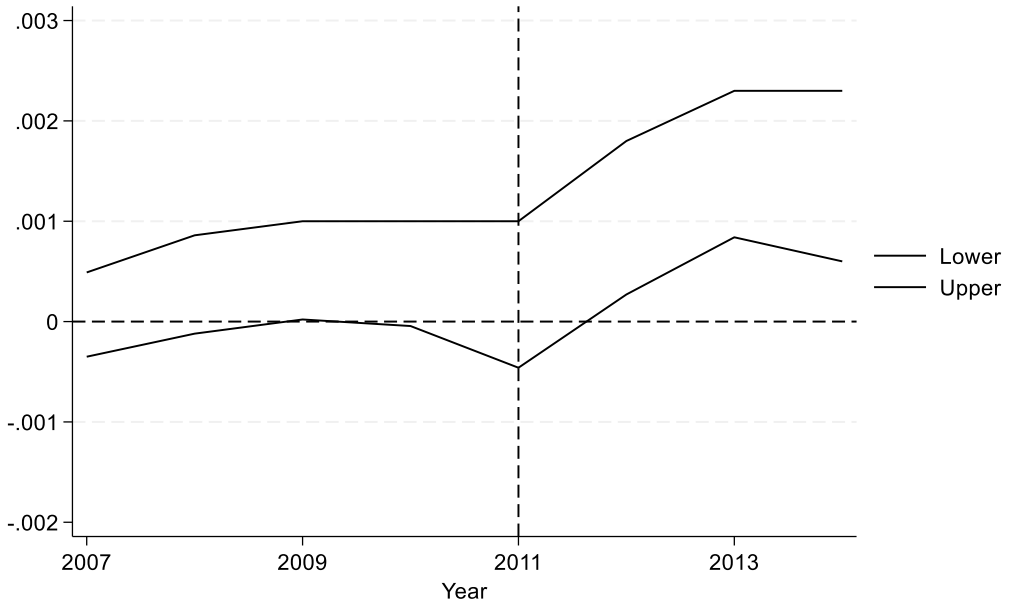
The results presented in Table 7 and Figure 3 show increases across all three tax outcome variables due to the reform. Over the 4-year post-reform period, the average treatment effects are 0.002 for $TOTAL\_TAX_{ft}$, 0.001 for $CONFTAX_{ft}$, and 0.008 for ETR_{ft}. These effects are all statistically significant at the 1% level, as confirmed by an F-test. The increase in total taxes paid equals 6% (0.002/0.033 = 0.06) or approximately 3,700 SEK per firm per year on average (62,215 × 0.06 = 3,733). However, as mentioned above, the matching process and the associated loss in observations mean that this estimate reflects only a within-sample effect of the reform.

Using the numbers for ETR_{ft}, which exhibits higher external validity, we obtain the following. The effective tax rate changes by 0.8 percentage points and the average profit after net financial expenses (which is the base the corporate tax is levied upon) in the pre-reform period equals 266,415 SEK per firm per year for the voluntarily audited firms. As such, the taxes paid increased by approximately 2,100 SEK (266,415 × 0.008 = 2,131) in voluntarily audited firms due to the reform. The increase in the effective tax rate of the voluntarily audited firms can be attributed to greater increases in both tax payments and profits for these firms than for their mandatorily audited counterparts after the implementation of the reform, with the increase in tax payments being slightly greater than the increase in profits.

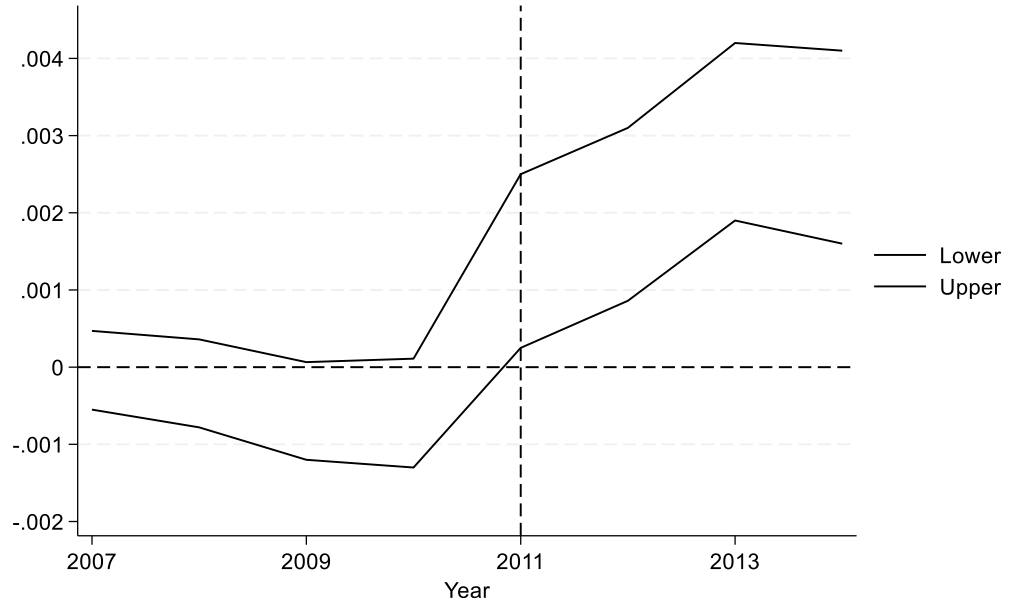
Table 7. Year-by-year treatment effect estimations.

Variable	TOTAL TAX	CONFTAX	ETR
VOL_AUDIT_f × Year 2007	-0.000	0.000	0.004
	(-0.15)	(0.33)	(1.34)
VOL_AUDIT_f × Year 2008	-0.000	0.000	0.000
	(-0.73)	(1.46)	(0.08)
VOL_AUDIT_f × Year 2009	-0.001*	0.001**	0.003
	(-1.75)	(2.04)	(0.93)
VOL_AUDIT_f × Year 2010	-0.001	0.001*	0.000
	(-1.63)	(1.80)	(0.14)
VOL_AUDIT_f × Year 2011	0.001**	0.000	0.006*
	(2.40)	(0.75)	(1.91)
VOL_AUDIT_f × Year 2012	0.002***	0.001***	0.009***
	(3.45)	(2.68)	(2.96)
VOL_AUDIT_f × Year 2013	0.003***	0.002***	0.009***
	(4.99)	(4.12)	(2.72)
VOL_AUDIT_f × Year 2014	0.003***	0.001***	0.009***
	(4.45)	(3.36)	(2.84)
Firm FE	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes
Control variables	Yes	Yes	Yes
P-value, F-test 2007–2010	0.20	0.08	0.45
Observations	41,328	66,762	268,596
R^2	0.11	0.19	0.18

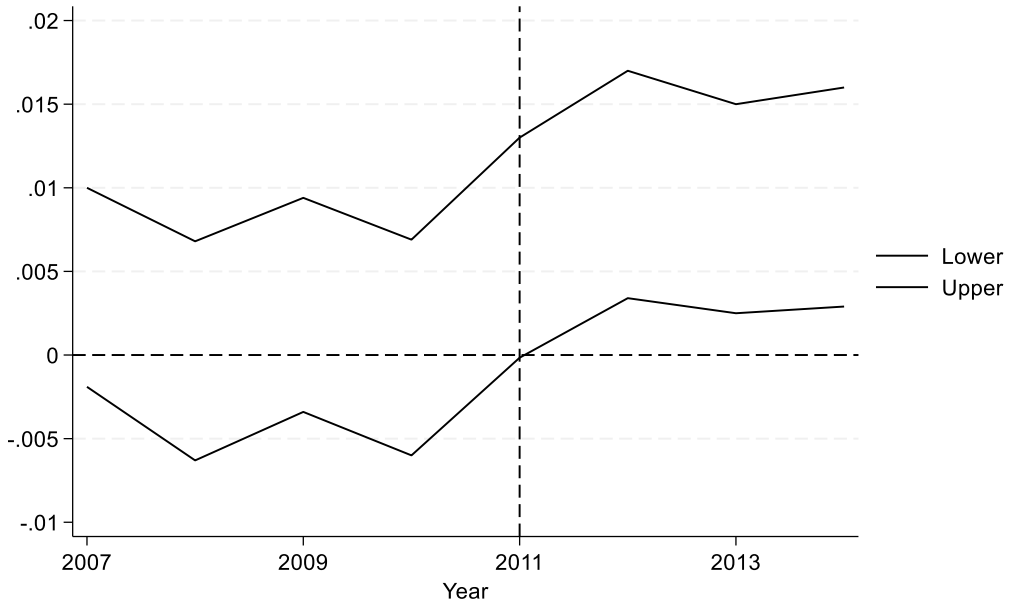
Notes: t-values are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. 2006 is used as the base year in the estimations. ETR is estimated on the full sample, as no matching is necessary to achieve parallel trends according to the F-test. Tax scaled by assets and conftax are estimated on CEM-matched samples using 50 intervals in the matching of the outcome variable in the pre-reform period.



a) TOTAL TAX



b) CONFTAX



c) ETR

Figure 3a-c. Year-by-year treatment effect estimates, 95% confidence intervals.

4.2 Placebo testing of the results

As mentioned in footnote 10, the annual treatment effect estimates serve as a placebo test across time. For the period before the actual reform, the effective tax rate, ETR_{ft}, which is the outcome variable with the highest external validity, shows no statistically significant deviations from 0 for any single year. For the period after the reform, we observe an adjustment in the effective tax rate during the first year. However, the estimated treatment effect stabilizes at similar levels for the following 3 years.

Additionally, placebo tests were conducted by manipulating firm sizes in the same manner as in Section 3.2, though now using the control group firms identified in the estimations in Section 4.1. If we have identified true reform effects, such a placebo test should not find any statistically significant treatment effects, as the analysis is conducted only for firms that have never been eligible for voluntary audits. The results from these placebo tests are presented in Table 8, where no statistically significant treatment effects were identified for any of our outcome variables— $TOTAL\_TAX_{ft}$, $CONFTAX_{ft}$, or ETR_{ft}.

Table 8: Placebo test manipulating the size of the reform thresholds.

	Placebo test size		
	TOTAL TAX	CONFTAX	ETR
$VOL\_AUDIT_f \times POST_t$	-0.000	-0.000	0.003
	(-0.43)	(-0.96)	(1.12)
Firm FE	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes
Control variables	Yes	Yes	Yes
Observations	13,591	22,377	71,365
R^2	0.12	0.15	0.17

Notes: t-values are reported in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01. Placebo estimations are performed on control group firms with mandatory audits.

5. Summary and discussion

The purpose of this study was to conduct a replication of Dong et al.'s (2023) study. To achieve this, we began by performing a reproduction replication study, as defined by Babin et al. (2021), of the primary empirical analysis presented by Dong et al. (2023) to verify whether we could replicate the original results. In this part, we also included placebo testing of the results from our replication analysis. Subsequently, we conducted what Babin et al. (2021) define as a differentiated replication, incorporating extensions of Dong et al.'s (2023) study. This part of the analysis deliberately modified the methods to investigate whether the original results would hold under more deliberate empirical models.

In the reproduction replication phase, we were able to replicate the results of Dong et al. (2023) with reasonable accuracy. Although the parameter estimates in the analysis differed in size, Dong et al.'s (2023) general conclusion held consistent, namely that the voluntary audit reform led to a reduction in tax payments for treated firms. However, placebo testing conducted on our data – which was not performed by Dong et al. (2023) – showed that the fundamental assumption in the DiD analysis, of parallel trends in the absence of treatment, was likely violated in the analysis.

We then employed a step-by-step method, beginning with an investigation of the parallel trend assumption on the original, pre-matching datasets. This analysis showed that for the effective tax rate, ETR_{ft}, the original data displayed parallel trends in the pre-reform period without requiring matching. However, for $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$, matching was necessary to achieve parallel trends. The results from estimations using datasets where pre-reform trends could be demonstrated as parallel for all outcome variables, show – contrary to the analysis of

Dong et al. (2023) and our replication of their results – that tax payments increased due to the voluntary audit reform.

These findings raise several interesting topics for further discussion. First, the current widespread use of conditional DiD estimation (Heckman et al., 1998), which involves matching to construct suitable control groups before DiD estimation, should not, in our opinion, be applied indiscriminately. The first-step matching used in this method often comes at a high cost in terms of the loss of observations and associated reduction in the external validity of the research—particularly when the stated purpose of the research is to make inferences about the entire population under study. We therefore suggest that the identification assumption of DiD, which posits that the trends of the outcome variables would have been parallel in the absence of treatment, should first be investigated using the complete datasets. If there is reasonable suspicion that the assumption is not valid in the complete dataset, researchers can then turn to matching and conditional DiD analysis.

Moreover, our results suggest that placebo testing should be employed more extensively in current empirical research. As highlighted in Section 3, a simple tabulation of 1-year differences in outcome variables between 2007 and 2010 for treatment and control group firms, accompanied by graphs displaying the development of the outcome variable, might lead to the incorrect conclusion that trends were parallel. However, placebo testing clearly revealed this not to be the case, as statistically significant treatment effects were observed where none should exist.

These methodological issues also have implications for the choice of outcome variables in final analyses and for policy recommendations. While we identified samples with parallel trends for all three outcome variables in the period leading up to the reform, the necessary matching procedure for $TOTAL\_TAX_{ft}$ and $CONFTAX_{ft}$ resulted in significant losses of observations. In contrast, ETR_{ft} did not experience such losses. Thus, the estimation of the treatment effects for ETR_{ft} is likely to better represent the population under study and thus serve as a stronger basis for policy recommendations.

Focusing on the reform's impact on the effective tax rate, ETR_{ft}, we find that voluntarily audited firms increased, rather than reduced, their tax payments due to the reform. It should also be noted that the treatment effects for ETR_{ft} are never negative and statistically significant at conventional levels, either in our reproduction replication study or in Dong et al. (2023). The results of our extended study regarding the treatment effect of voluntary auditing on the effective tax rate instead show that the reform increased tax payments. Our best estimate

suggests that this increase amounts to, on average, 2,100 SEK per firm per year. The increase in the effective tax rate for voluntarily audited firms can be attributed to greater increases in both tax payments and profits for these firms compared to their mandatorily audited counterparts after the reform's implementation, with the increase in tax payments slightly greater than the increase in profits.

The finding that voluntarily audited firms experience greater profit increases compared to mandatorily audited firms is partially attributable to selection effects. Specifically, successful firms with growth ambitions opt for audits, as they anticipate surpassing regulatory thresholds (Ojala et al., 2016). Another contributing factor is that voluntarily audited firms often benefit from a lower cost of capital for investments compared to their mandatorily audited competitors (Dedman et al., 2014; Huq et al., 2022; Kim et al., 2011). Additionally, the relatively greater increase in tax payments observed among voluntarily audited firms as profits increase may result from these firms having exhausted legal opportunities for tax-reducing adjustments in their financial statements. Exploring whether this is the case, and to what extent, is a complex task due to the intricacies of the Swedish tax code, making it a compelling question for future research.

Acknowledgements

The authors would like to thank the Confederation of Swedish Enterprise (Rudholm) and Jan Wallander and Tom Hedelius Foundation (P23-0312, Hartwig; P23-0200, Huq) for financial support. The Confederation of Swedish Enterprise or the Jan Wallander and Tom Hedelius Foundation were not involved at any stage in the study design, data collection, data analysis, or decision to submit the paper for publication. The authors also wish to thank Sven-Olov Daunfeldt, Jonas Frycklund and Claes Norberg, as well as participants at seminars at the Confederation of Swedish Enterprise on 11 November 2024, and at Dalarna University on 12 February 2025, for their valuable comments and suggestions.

References

Abadie, A. (2005). Semiparametric difference-in-difference estimators. *Review of Economic Studies*, *72*(1), 1–19.

Accountancy Europe, 2021, Audit exemption thresholds in Europe. Available at: https://www.accountancyeurope.eu/wp-content/uploads/2022/12/Audit-exemption-thresholds-in-Europe.pdf

Alduraywish, Y. (2023). Do audited firms have better access to credit?: Evidence from emerging countries. *Cogent Business & Management*, *10*(2), 2195985.

Arcidiacono, P., Ellickson, P. B., Mela, C. F., & Singleton, J. D. (2020). The competitive effects of entry: Evidence from supercenter expansion. *American Economic Journal: Applied Economics*, *12*(3), 175–206.

Audretsch, D. B. (1995). Innovation, growth and survival. *International Journal of Industrial Organization*, *13*(4), 441–457.

Audretsch, D.B., Klomp, L., Santarelli, E., & Thurik, A.R. (2004). Gibrat's law: Are the services different? *Review of Industrial Organization*, *24*(3), 301–324.

Babin, B. J., Ortinau, D. J., Herrmann, J. L., & Lopez, C., 2021. Science is about corroborating empirical evidence, even in academic business research journals. *Journal of Business Research*, *126*, 504–511.

Badertscher, B. A., Katz, S. P., Rego, S. O., & Wilson, R. J. (2019). Conforming tax avoidance and capital market pressure. *The Accounting Review*, *94*(6), 1–30.

Blackwell, M., Iacus, S. M., King, G., & Porro, G. (2009). CEM: Coarsened exact matching in Stata. *The Stata Journal*, *9*(4), 524–546.

Clatworthy, M. A., & Peel, M. J. (2013). The impact of voluntary audit and governance characteristics on accounting errors in private companies. *Journal of Accounting and public policy*, *32*(3), 1–25.

Daunfeldt, S-O., Lang, Å., Macuchova, Z., & Rudholm, N. (2013) Firm growth in the Swedish retail and wholesale industries. *The Service Industries Journal*, *33*(12), 1193–1205.

Daunfeldt, S-O., & Rudholm, N. (2021). Den slopade revisionsplikten för små bolag: En granskning av granskarnas rapporter. HFI Forskningsrapport 2021:03.

Dedman, E. & Kausar, A. (2012). The impact of voluntary audit on credit ratings: Evidence from UK private firms. *Accounting and Business Research*, *42*(4), 397–418.

Dedman, E., Kausar, A., & Lennox, C. (2014). The demand for audit in private firms: recent large-sample evidence from the UK. *European Accounting Review*, *23*(1), 1-23.

Downing, J. & Langli, J. C. (2019). Audit exemptions and compliance with tax and accounting regulations. *Accounting and Business Research*, 49(1), 28–67.

Dong, T., Tylaite, M., & Wilson, R. (2023). Voluntary vs. mandatory: the role of auditing in constraining corporate tax avoidance in small private firms. *Accounting and Business Research*, *53*(7), 723–755.

European Economic Community, 1978. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31978L0660

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica*, *66*(5), 1017–1098.

Huq, A.M., Daunfeldt, S.O., Hartwig, F., & Rudholm, N. (2021). Free to choose: do voluntary audit reforms increase employment growth? *International Journal of the Economics of Business*, *28*(1), 163–178.

Huq, A.M., Hartwig, F., & Rudholm, N. (2022). Do audited firms have a lower cost of debt? *International Journal of Disclosure and Governance*, *19*(2),153–175.

Iacus, S. M., King, G., & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, *106*(493), 345–361.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*(1), 1–24.

Keasey, K., Watson, R., & Wynarczyk, P. (1988). The small company audit qualification: a preliminary investigation. *Accounting and Business Research*, *18*(72), 323–334.

Kim, J.-B., Simunic, D. A., Stein, M. T., & Yi, C. H. (2011). Voluntary audits and the cost of debt capital for privately held firms: Korean evidence. *Contemporary Accounting Research*, *28*(2), 585–615.

Koren, J., Kosi, U., & Valentincic, A. (2014). Does financial statement audit reduce the cost of debt of private firms? *SSRN Electronic Journal*, 2373987.

Ojala, H., Collis, J., Kinnunen, J., Niemi, L., & Troberg, P. (2016). The demand for voluntary audit in micro-companies: Evidence from Finland. *International Journal of Auditing*, *20*, 267–277.

Svanström, T. & Sundgren, S. (2012). The demand for non-audit services and auditor-client relationships: Evidence from Swedish small and medium-sized enterprises. *International Journal of Auditing*, *16*(1), 54–78.

Swedish National Audit Office [Riksrevisionen] (2017). Avskaffandet av revisionsplikten för små bolag: En reform som kostar mer än det smakar. RIR 2017:35. Available at: https://www.riksrevisionen.se/granskningar/granskningsrapporter/2017/avskaffandet-av-revisionsplikten-for-sma-aktiebolag---en-reform-som-kostar-mer-an-den-smakar.html

Swedish Economic Crime Authority [Ekobrottsmyndigheten] (2016). Effekter på den ekonomiska brottsligheten efter avskaffandet av revisionsplikten för mindre aktiebolag. Rapport, juni 2016. Available at: https://www.ekobrottsmyndigheten.se/wp-content/uploads/2021/06/pm-2016-06-01-rapport-slutversion.pdf

Swedish Economic Crime Authority [Ekobrottsmyndigheten] (2020). Fördjupad studie av effekter på den ekonomiska brottsligheten efter avskaffande av revisionsplikten för mindre aktiebolag. Rapport, april 2020. Available at: https://www.ekobrottsmyndigheten.se/wp-content/uploads/2021/06/fordjupad-studie-av-effekter-pa-den-ekonomiska-brottsligheten-efter-avskaffande-av-revisionsplikten-for-mindre-bolag.pdf

Swedish Government, 2010. Regeringens proposition 2009/10:204, En frivillig revision (2009/10:204). Available at: http://www.regeringen.se/contentassets/b7693839f66f49079138ebad54098762/en-frivillig-revision-prop.-200910204

Swedish Government, 2010. Svensk författningssamling, SFS 2010:834. Available at: https://www.notisum.se/rnp/sls/sfs/20100834.pdf

Sveriges Riksdag, 2005. Aktiebolagslag 2005:551. Available at: http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/aktiebolagslag-2005551_sfs-2005-551

Öhman, P. & Wallerstedt, E. (2012). Audit regulation and the development of the auditing profession: The case of Sweden. *Accounting History*, *17*(2), 241–257.

Online supplemental material

Replicating and extending 'Voluntary vs mandatory: the role of auditing in constraining corporate tax avoidance in small private firms'

Fredrik Hartwig, Asif M Huq, Niklas Rudholm

Appendix A1. Data collection and editing.

In Sweden, all limited liability firms are required to submit annual reports containing financial statement data to the Swedish Companies Registration Office (SCRO). The data used for this replication has been compiled by the Swedish House of Finance National Research Data Center and licenced to the University of Gävle for research. The dataset used is thus the property of the Swedish House of Finance National Research Data Center. Accordingly, we cannot provide open access to this dataset or deposit it in an open online repository. However, this dataset, or similar datasets based on data from the SCRO, are available through several sources, and most Swedish universities have access to some form of these data. Any such dataset can be used when attempting to replicate our study.

Our starting sample consists of all independent Swedish private limited firms active at any point from 2006 to 2020. According to the database, firms are active when the variable 'ser_aktiv' is equal to '1'. Subsidiaries and parents of groups are excluded, using the variable 'knc_kncfall'. We then exclude firm-year observations with missing information on key variables: profit or loss after net financial income or expenses [rr12_resefin], taxation [rr14_skatter], total assets [br09_tillgsu], operating profit or loss [rr07_rorresul], net sales [rr01_ntoms], machinery and equipment [br03_maskiner], trade accounts payable [br13a_ksklev], cash and bank balance [br07b_kabasu], and retained earnings [br10e_balres]. Furthermore, firms with missing information on audit status and board information are excluded. Next, we exclude firm-year observations with missing information on variables necessary for the replication of Dong et al. (2023) and industry-year combinations with fewer than 5 observations. In Table A1.1, we report the number of firms and firm-year observations during the different steps of the sample construction.

Table A1.1. Sample construction.

Details	Unique firms	Firm-years
Starting sample of active independent firms, 2006–2016	415,356	2,354,105
Missing on raw-variables, audit status, and board information	(20,251)	(129,444)
Missing during constructed variables and industry-year with fewer than 5 observations	(110,344)	(573,017)
All firms for the years 2006–2016	**284,761**	**1,651,644**

In Table A1.2, we report the sample distributions for the Dong et al. (2023) replication and the two placebo tests across time. We begin with all firm-year observations for the specified time periods, namely 2006–2014 for the Dong et al. (2023) replication and 2006–2010 and 2012–2016 for the two 5-year placebo tests across time for the placebo reform years 2008 and 2014, respectively. Following Dong et al. (2023), we only include firms active throughout the whole period (i.e., 9 years for the replication and 5 years for our placebo tests). Next, firms with stable voluntary or mandatory audits throughout the 9 or 5 years are retained. Finally, the matching procedure stipulated in Dong et al. (2023) is used to construct the matched sample for the replication and two placebo tests across time. For the Dong et al. (2023) replication, this results in a final sample of 7,686 firms (3,843 firms in each group) followed over a 9-year period, resulting in a total of 69,174 firm-year observations.

Table A1.2. Sample distribution: Dong et al. (2023) replication and time placebo tests.

Details	Dong et al. (2023) replication 2006–2014		Placebo reform [2008] 2006–2010		Placebo reform [2014] 2012–2016	
	Unique firms	Firm-years	Unique firms	Firm-years	Unique firms	Firm-years
All firms	242,612	1,283,853	177,211	660,327	237,363	844,832
Active throughout	68,271	614,439	90,737	453,685	108,279	541,395
Stable voluntary audit	18,698	168,282	61,776	308,880	26,485	132,425
Stable mandatory audit	11,146	100,314	17,252	86,260	20,410	102,050
Matched sample	**7,686**	**69,174**	**16,388**	**81,940**	**19,626**	**98,130**

Finally, in Table A1.3, we report the sample distributions for the size placebo test. We begin with the matched sample used in the replication of Dong et al. (2023). We then exclude firms with stable voluntary audits (i.e., only control group firms with stable mandatory audits are retained). Then, we set the following placebo conditions for constituting a 'treated' firm:

i. The average number of employees for the previous consecutive fiscal year amounts to more than three but less than six.

ii. Reported total assets for the previous consecutive fiscal year amounts to more than 1.5 but less than 3.0 million SEK.

iii. Reported net sales for the previous consecutive fiscal year amounts to more than 3.0 but less than 6.0 million SEK.

Lastly, for the control group, we use all firms in the dataset exceeding these thresholds with up to 10 employees. Here, we determine the treatment indicator based on exceeding the limits by 1 year, instead of 2 years as stipulated in the original reform, and allow firms to move in and out of the placebo treatment group. We do this to avoid losing an excessive number of observations when using lagged variables and stable placebo treatment throughout the 9-year period. Additionally, since there is no risk of selection into treatment bias when investigating a fictitious reform as in our size placebo test.

Table A1.3. Sample distribution: Size placebo test.

Details						**2006–2014**			
						Unique firms		**Firm-years**	
Matched sample for Dong et al. (2023) replication						7,686		69,174	
Stable mandatory audit						3,843		34,587	
Firms with fewer than 10 employees						3,537		29,691	
Year	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
Placebo treated	1,485	1,315	1,284	1,406	1,322	1,212	1,224	1,257	1,151
Placebo control	1,879	1,993	2,008	1,907	1,989	2,077	2,059	2,012	2,111

Appendix A2. Year-by-year estimations.

In this section, we present the results from estimations where we begin with a dataset and the basic data editing necessary, including winsorizing the variables, but before any type of matching. On these data, we then implement a more rigorous test of the parallel trend assumption than descriptive graphs, following Arcidiacono et al. (2020), modifying our estimating equation to include year-by-year treatment effect parameters. More formally, using the same control variables and fixed effects in the replication of Dong et al. (2023), the following year-by-year DiD model is estimated:

$$TAX_{ft} = \alpha + \sum_{t=2007}^{2014} \beta_t \left(VOL_{AUDIT_f} \times YEAR_t\right) + \gamma' \mathbf{X} + \mathrm{FIRM_{FE}}$$

$$+ \mathrm{IND} \times \mathrm{YEAR_{FE}} + \varepsilon_{ft} \qquad (2)$$

If the parallel trend assumption holds, the treatment effect estimates for the years leading up to the reform will be statistically insignificant and close to 0. This is tested both by individual t-tests of the treatment effect parameter estimates for the years leading up to the reform and by an F-test investigating whether the treatment effect parameter estimates for the years leading up to the reform are jointly significantly different from 0 at the 5% level. If they are, the parallel trend assumption is invalid, and no causal interpretation of the post-reform treatment effect estimates can then be made.

The results presented in Table A2.1 show that the parallel trend assumption is satisfied on the complete dataset for the effective tax rate variable, though not for the variables Taxes scaled by assets or Dong conftax. Accordingly, we use coarsened exact matching (CEM; Blackwell et al., 2009; Iacus et al., 2011, 2012) on the outcome variables Taxes scaled by assets and Dong conftax during the pre-reform years to find control group firms with a similar development of these variables in the years leading up to the reform. In the first round of matching, we use Sturges's rule to determine the size of the intervals in the matching.

In Table A2.2 below, for Tax scaled by assets, the F-test cannot verify that the treatment effects in the years leading up to the reform are different from 0, suggesting that pre-treatment trends could be parallel. However, this is due to unstable treatment effect estimates in these years, with some positive and some negative and statistically significant effects. The parallel trend assumption for DiD is thus not fulfilled, even if the F-test for taxes scaled with assets suggests so (while individual t-tests for the different years do not). A similar, but less apparent, pattern emerges for the Dong conftax variable.

To address this problem, we increase the number of intervals in the CEM from approximately 20 (as suggested by Sturges's rule) to 50. The results from estimations using these data for the Tax scaled by assets and Dong conftax variables are presented in Table 7 in the main text, where the t- and F-tests indicate that the pre-treatment trends are (close to) parallel. All estimations (TOTAL TAX, CONFTAX, and ETR) begin with 268,596 observations, but due to the two rounds of matching necessary to achieve parallel trends for TOTAL TAX and CONFTAX, the final estimations result in 41,328 observations for TOTAL TAX and 66,762 observations for CONFTAX.

Table A2.1. Year-by-year treatment effect estimations, before matching.

Variable	TOTAL TAX	CONFTAX	ETR
VOL_AUDIT_f × Year 2007	0.00070*	0.00086***	0.0041
	(1.68)	(2.69)	(1.34)
VOL_AUDIT_f × Year 2008	0.0022***	0.0025***	0.00027
	(4.76)	(7.32)	(0.08)
VOL_AUDIT_f × Year 2009	0.0021***	0.0024***	0.0030
	(4.58)	(6.96)	(0.93)
VOL_AUDIT_f × Year 2010	0.0022***	0.0024***	0.00045
	(4.52)	(6.12)	(0.14)
VOL_AUDIT_f × Year 2011	0.0022***	0.0027***	0.0063*
	(4.51)	(7.29)	(1.91)
VOL_AUDIT_f × Year 2012	0.0021***	0.0031***	0.0099***
	(4.42)	(8.45)	(2.96)
VOL_AUDIT_f × Year 2013	0.0028***	0.0031***	0.0090***
	(5.89)	(8.51)	(2.72)
VOL_AUDIT_f × Year 2014	0.0026***	0.0030***	0.0094***
	(5.21)	(7.84)	(2.84)
Firm FE	Yes	Yes	Yes
Industry-year FE	Yes	Yes	Yes
Control variables	Yes	Yes	Yes
P-value, F-test 2007–2010	0.00	0.00	0.45
Observations	268,596	268,596	268,596
R^2	0.13	0.04	0.12

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. t-values are in parenthesis. 2006 is used as the base year in the estimations.

Table A2.2. Year-by-year treatment effect estimations, after initial matching using Sturges's rule to determine the number of intervals in the matching.

Variable	Tax scaled by assets	Dong conftax
VOL_AUDIT_f × Year 2007	0.0012***	-0.000092
	(2.94)	(-0.30)
VOL_AUDIT_f × Year 2008	-0.000055	0.00067**
	(-0.15)	(2.26)
VOL_AUDIT_f × Year 2009	0.0010**	0.00064*
	(2.41)	(1.95)
VOL_AUDIT_f × Year 2010	-0.00079**	0.00070**
	(-1.99)	(2.22)
VOL_AUDIT_f × Year 2011	0.0012***	0.00063*
	(2.62)	(1.73)
VOL_AUDIT_f × Year 2012	0.0016***	0.0014***
	(3.54)	(3.78)
VOL_AUDIT_f × Year 2013	0.0025***	0.0015***
	(5.38)	(4.05)
VOL_AUDIT_f × Year 2014	0.0023***	0.0015***
	(4.68)	(4.00)
Firm FE	Yes	Yes
Industry-year FE	Yes	Yes
Control variables	Yes	Yes
P-value, F-test 2007–2010	0.29	0.05
Observations	132,408	144,702
R^2	0.13	0.07

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. t-values are in parenthesis. 2006 is used as the base year in the estimations.

Appendix A3. Trends in the outcome variables ETR and CONFTAX in the final datasets used for the estimations presented in Table 7.

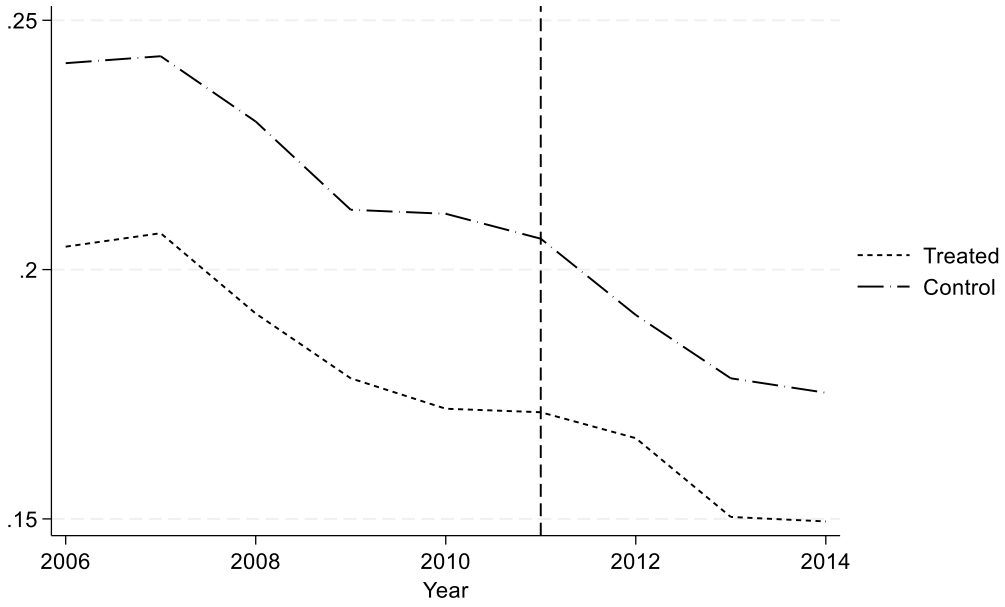


Figure A3.1. Effective tax rates (ETR) in voluntary (treated) and mandatorily (control) audited firms, before matching.

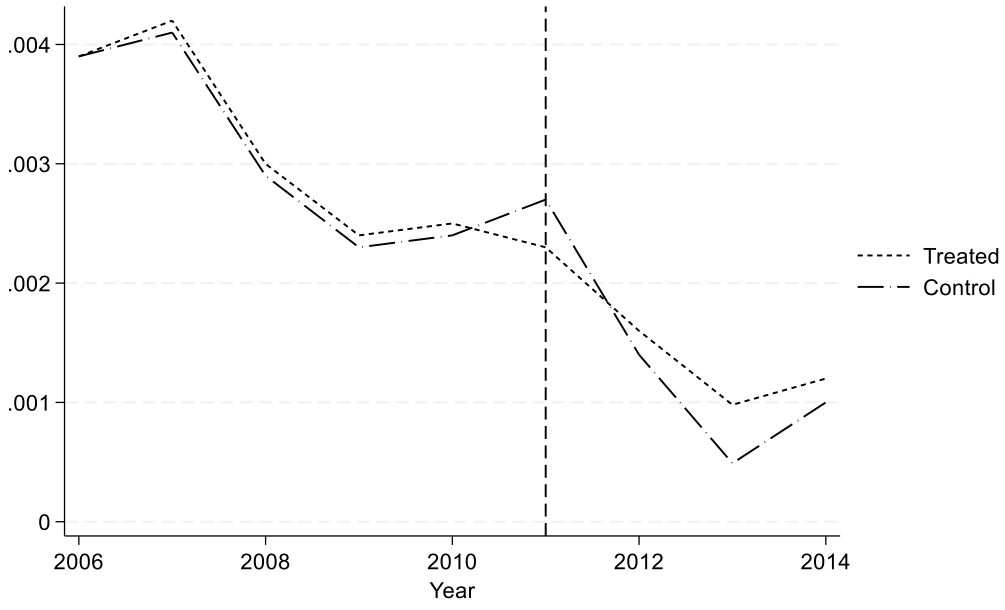


Figure A3.2. Dong conftax (CONFTAX) in voluntary (treated) and mandatorily (control) audited firms, after CEM matching using 50 intervals.