

Klimsa, Drahomir; Ullmann, Robert

Article — Published Version

Threshold-dependent tax enforcement and the size distribution of firms: evidence from Germany

International Tax and Public Finance

Provided in Cooperation with:

Springer Nature

Suggested Citation: Klimsa, Drahomir; Ullmann, Robert (2022) : Threshold-dependent tax enforcement and the size distribution of firms: evidence from Germany, International Tax and Public Finance, ISSN 1573-6970, Springer US, New York, NY, Vol. 30, Iss. 4, pp. 1002-1035, <https://doi.org/10.1007/s10797-022-09732-2>

This Version is available at:

<https://hdl.handle.net/10419/309509>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Threshold-dependent tax enforcement and the size distribution of firms: evidence from Germany

Drahomir Klimsa¹ · Robert Ullmann¹

Accepted: 23 February 2022 / Published online: 22 April 2022
© The Author(s) 2022

Abstract

This paper investigates firms' responses to threshold-dependent intensity of tax enforcement. We use administrative tax return data over the entire population of German firms and exploit industry variation in firm size thresholds applied by the tax administration. In our setting, each threshold marks a considerable spike in audit intensity and hence should create strong incentives to bunch below the threshold. However, we find no such effect in our large sample analysis. We attribute this empirical observation to optimization costs, particularly to the costs associated with the operational implementation of size management and to information costs. Our paper adds to the emerging field of studies on potential distortions created by threshold-dependent firm regulation. The findings are also relevant for policymakers, as they suggest that the specific design of threshold-dependent policies might allow governments to increase the efficiency of tax audits without distorting the firm size distribution.

Keywords Tax enforcement · Size-dependent regulation · Bunching · Administrative data · Germany

JEL Classification H26 · H32 · K42

1 Introduction

Large firms are subject to higher audit intensity from tax administrations than small firms (Bachas et al., 2019) because governments segment taxpayers by firm size in order to increase the efficiency of tax audits. Naturally, a tax audit is costly for the

✉ Robert Ullmann
robert.ullmann@wiwi.uni-augsburg.de

Drahomir Klimsa
drahomir.klimsa@wiwi.uni-augsburg.de

¹ University of Augsburg, Chair of Business Taxation, Universitätsstr. 16, 86159 Augsburg, Germany

firm, as the handling of the auditor creates compliance costs and any tax audit creates a nonzero probability of additional tax claims, interest payments and penalty fees. Hence, firms have incentives to avoid greater audit intensity. When audit intensity depends on firm size, firms have reason to strategically bunch below firm size thresholds (FSTs) through size management. Respective FSTs are often made publicly available by tax administrations. However, size management distorts the firm size distribution and has negative effects on welfare. Specifically, due to the firms' costs of size management, size management results in a deadweight loss, reduces firm's future economic performance and, consequently, overall economic growth and also decreases aggregate productivity because of inefficient resource allocation.

Despite the adverse effects that result from size management, research on this subject is scarce. To our knowledge, only two studies analyze how firms respond to threshold-dependent tax enforcement on the microlevel. First, Almunia and Lopez-Rodriguez (2018) find significant downward size management by Spanish firms and present evidence that underreporting of revenue is the key channel for this phenomenon in their setting. Second, Tennant and Tracey (2019) examine a threshold-dependent policy in Jamaica. In contradiction to the results by Almunia and Lopez-Rodriguez (2018), Tennant and Tracey (2019) find no size management around the FST.

However, prior research has shown size management in many areas of taxation other than tax enforcement. For instance, Hoopes et al. (2018) analyze the responses of Australian firms to the threshold-dependent intensity of tax return disclosure. They find that firms manage their size to avoid disclosure. Further research has shown that size management at FSTs is related to kinks in corporate income tax (CIT) (Brockmeyer, 2014; Devereux et al., 2014), CIT notches (Bachas and Soto, 2021), CIT benefits (Hosono et al., 2018) and special CIT regimes for SMEs (Agostini et al., 2018), minimum CIT schemes (Best et al., 2015) and exemptions in value added tax (VAT) (Harju et al., 2016; Liu et al., 2019; Onji, 2009). Moreover, size management has also been found in an array of nontax areas, e.g., mandatory IFRS reporting (Asatryan and Peichl, 2017), financial audit and disclosure requirements (Bernard et al., 2018) and labor regulation (Garicano et al., 2016).

We use administrative microlevel tax return data to study size management for the entire population of German firms. These firms face threshold-dependent discontinuities in audit intensity. Specifically, the German tax administration segments firms into four size classes based on FSTs: very small (VS-class), small (S-class), medium (M-class) and large (L-class). Firms are assigned to a particular size class if their size exceeds either the respective FST for profit or for revenue (or both). The FSTs vary between industries (and increase continuously over time). Audit intensity between size classes varies most notably in terms of audit probability. For instance, in 2010, 21.1% of firms assigned to the L-class were audited as opposed to only 6.9% of firms in the M-class. In the S-class and the VS-class, the audit probabilities were only 3.5% and 1%, respectively (German Federal Ministry of Finance, 2011). Both the FSTs and the corresponding audit probabilities are regularly published online on the website of the Federal Ministry of Finance and in the Federal Gazette. In addition to audit probability, audit intensity between size classes also varies in terms of audit

quality. First, administrative regulation dictates that for L-class firms, the audit must be consecutive, i.e., once an audit occurs, it must cover all years that were not covered by the previous audit for that firm. In contrast, for M-class, S-class and VS-class firms, the audit period cannot exceed three calendar years. Second, the skill level of the auditor and the specialization level of audit teams increase systematically with the size class.

Our results imply an absence of size management in our data. Although, naturally, the null of no size management cannot be proven, a type II error is unlikely in our setting. First, our dataset is large, with approximately 2.7 million firms included. This substantially reduces the probability of making a type II error, even in our most granular subsample analysis, in which we search for size management in individual industries. Furthermore, as we rely on administrative data, we arguably face negligible measurement error and no selection bias. Finally, the results do not seem to be driven by our specific empirical strategy, as we obtain structurally equivalent results when applying an array of alternative tests.

We make a contribution to the emerging field of studies on potential distortions created by threshold-dependent firm regulation in showing that firms in our setting do not react to FSTs by size management despite strong incentives to the contrary. We posit that the absence of size management results from optimization costs in the form of adjustment costs and information costs. The results we find for Germany contradict the results found by Almunia and Lopez-Rodriguez (2018) for Spain despite both countries being relatively similar in relevant drivers of optimization costs. Specifically, the two are similarly developed countries located in Western Europe, do not differ substantially in terms of the level of trust in public institutions and have similar tax rates. Despite these similarities, Germany and Spain differ in the specific design of their threshold-dependent enforcement regime. We argue that Germany's specific implementation of multiple criteria for segmentation, multiple size classes, regular adjustments of FSTs and industry-specific FSTs increase optimization costs and, hence, can inhibit tax-induced size management. Moreover, the results by Tennant and Tracey (2019) on firms in Jamaica, where FSTs are based on a combination of taxes paid and revenue, indicate that a more multilayered threshold-dependent policy improves firms' tax compliance as measured by both reported profitability and effective tax rates without causing tax-induced size management.

Overall, this field of research is relevant for policymakers, as the results suggest that the specific design of threshold-dependent policies might allow governments to increase the efficiency of tax audits while not distorting the firm size distribution and, hence, avoid the negative effects of size management on welfare.

The remainder of this paper is organized as follows. Section 2 outlines the effects of threshold-dependent tax enforcement and the rationale of tax-induced size management. Section 3 provides information on the German tax enforcement regime. Section 4 develops our hypotheses, and Sect. 5 describes the empirical strategy. Section 6 presents information on data and on sample selection. Section 7 provides the main empirical results as well as a discussion of them. Section 8 contains robustness tests, and Sect. 9 concludes.

2 Literature and theoretical discussion

2.1 Effects of threshold-dependent tax enforcement

Governments worldwide focus their audit resources on large business taxpayers. Specifically, approximately 85% of the world's 60 largest economies segment firms into size classes based on FSTs and apply higher audit intensities to firms in the upper size classes (OECD, 2015). The major reason for the establishment of threshold-dependent policies is that they are believed to improve the efficiency of tax audits and preserve audit budgets. Furthermore, these policies aim to secure the integrity of the tax system, as larger taxpayers bear higher compliance risks than do smaller taxpayers (OECD, 2017). Operationally, most tax administrations differentiate between two size classes, and the FSTs applied to segment taxpayers are usually based on revenue, profit, total assets, taxes paid, the number of employees or a combination of these factors. The majority of tax administrations make respective FSTs publicly available.¹

On average, in countries that rely on threshold-dependent enforcement, firms exposed to the highest level of audit intensity provide 35–50% of the total tax revenue collected while representing less than 10% of all active firms (OECD, 2017). Focusing audit resources on a relatively small number of large firms appears efficient. There is also ample empirical evidence suggesting that tax compliance increases with audit intensity (Alm, 2019).² However, as shown by Alm et al. (2009), higher audit intensity has a positive impact on compliance only if taxpayers are well informed that they face a higher audit intensity. Hence, publicly available information about FSTs jointly with the respective historical audit rates, as an indicator for audit probability, can have positive effects on compliance.

As tax audits usually cause substantial costs for the audited firm, public information about FST levels may also trigger a size management response. Specifically, if firms above an FST face a significantly higher audit intensity than firms located below this FST, threshold-dependent enforcement policies create incentives to manage size below the FST. However, size management distorts the firm size distribution and has negative effects on welfare for several reasons. First, the firms' costs of size management represent an allocative inefficiency and thus result in a deadweight loss (Almunia and Lopez-Rodriguez, 2018). Second, as firms that manage their size in one period will also manage their size in future periods, size management has negative effects on firms' future economic performance (Roychowdhury, 2006) and, consequently, overall economic growth. Third and finally, size management also results in inefficient resource allocation and decreases aggregate productivity (Harju et al., 2016).

¹ For an overview of the criteria applied worldwide, see OECD (2015) and OECD (2017).

² For instance, see Hoopes et al. (2012) for recent evidence on public firms in the USA facing a higher IRS audit probability undertaking less aggressive tax positions compared to those facing lower audit probabilities.

Despite these negative effects on firms, research on this subject is scarce. On the macroeconomic level, Vehorn (2011) analyzes the impact of threshold-dependent tax enforcement policies in developing economies. The results show that 43% of countries experienced a decline in tax revenue (standardized by GDP) after the implementation of such policies, indicating adverse effects of threshold-dependent enforcement policies. On the microeconomic level, two studies analyze how firms respond to threshold-dependent tax enforcement. Both studies specifically investigate so-called large taxpayer units (LTUs), which are responsible for monitoring larger taxpayers. Firms are selected for LTU treatment when their size exceeds specific FSTs. First, Almunia and Lopez-Rodriguez (2018) find significant downward size management by Spanish firms at the revenue-based FST. Their results indicate that size management in their setting is predominantly conducted by underreporting rather than decreasing real activity. The results also indicate that the extent of tax-induced size management varies between industries conditional on the traceability of transactions due to third-party reporting. Traceability naturally determines the effectiveness of tax audits. Second, Tennant and Tracey (2019) examine an LTU policy in Jamaica, where FSTs are based on a combination of taxes paid and revenue. Their results indicate that LTU treatment significantly improves firms' tax compliance as measured by both reported profitability and effective tax rates. Contrary to the results by Almunia and Lopez-Rodriguez (2018), Tennant and Tracey (2019) find no size management at the FSTs. Overall, despite the widespread adoption of threshold-dependent enforcement regimes, the effects of such policies remain unclear.

2.2 Tax-induced size management

2.2.1 Rationale

We define tax-induced size management as any activity undertaken to manage firm size below an FST in order to reduce the firm's audit intensity, regardless of whether this activity is legal or illegal. Consistent with prior literature on notches in the tax system, e.g., Kanbur and Keen (2014), we argue that three nonmutually exclusive groups of size management strategies exist. First, firms can genuinely reduce their size by decreasing their real activity (also referred to as real production response). Second, firms can report a smaller size by using available discretion in accounting rules. For instance, firms can defer the recognition of revenue, create accruals or use special depreciations. Alternatively, firms can also split their operations into two or more individual legal entities (also referred to as tax-motivated splitting by Slemrod (2016)). Third, firms can report a smaller size by misreporting, e.g., by underreporting revenue or overreporting the cost of goods sold.

Regardless of the specific size management strategy, profit-maximizing firms engage in size management only as far as the benefits of size management exceed the resulting costs of size management (hereinafter referred to as optimization costs). The most notable benefit of size management is the decrease in expected costs from audits (hereinafter referred to as expected firm audit costs) when comparing the two

scenarios of firms just below and just above the FST. Consequently, if optimization costs exceed the decrease in expected firm audit costs around the FST, the threshold-dependent enforcement regime is not expected to distort the firm size distribution.

2.2.2 Expected firm audit costs

Expected firm audit costs can be defined as the costs that arise once a firm is audited (hereinafter referred to as conditional firm audit costs) multiplied by the probability that an audit occurs. Conditional firm audit costs represent a part of firms' total tax costs and consist of additional tax claims, interest payments, penalty fees and compliance costs.³ The first three elements are naturally conditional on detection and vary substantially in the cross section. As an example, variation between industries is conditional on the traceability of transactions under third-party reporting and hence conditional on the expected effectiveness of tax audits (Almunia and Lopez-Rodriguez, 2018). In contrast, considering the last element, compliance costs occur even if a firm is fully compliant. Compliance costs include costs of tax consulting services and administrative costs, i.e., the costs of employee resources allocated to the audit.⁴

2.2.3 Optimization costs

Optimization costs in the context of tax enforcement can be divided into adjustment costs and information costs. Whereas adjustment costs refer to the costs of operationally implementing size management, e.g., resource costs and opportunity costs of size management (Almunia and Lopez-Rodriguez, 2018), information costs result from gathering relevant information on the tax system, particularly on the threshold-dependent enforcement regime.

Adjustment costs are conditional on the criteria applied for segmentation. Specifically, size management in general is relatively difficult, as firms face uncertainty with respect to business outcomes throughout the year. However, while profit can often be adjusted through additional expenditures at the "last minute" when uncertainty declines by the end of the year (Asatryan et al., 2018), revenue, for example, is much more difficult to adjust.⁵ Correspondingly, revenue is applied for segmentation in approximately 70% of the countries that rely on threshold-dependent enforcement (OECD, 2017). Additionally, if multiple criteria have to be taken into account, size management becomes considerably more difficult and more time-consuming,

³ Recent research shows that besides causing costs, tax audits may also have positive effects for firms. Specifically, Guedhami and Pittman (2008) show that a higher audit probability reduces the costs of debt financing, and Gallemore and Jacob (2020) show that a higher audit probability increases commercial bank lending to firms. In general, however, it can be assumed that the costs of audits exceed potential benefits.

⁴ Firms worldwide spend approximately 25 hours complying with the requirements of an auditor and spend almost 11 weeks going through several rounds of interactions with the auditor according to The World Bank (2017).

⁵ Note that Bernard et al., (2018) found significant size management at FSTs related to financial audit and disclosure requirements in terms of total assets and the number of employees but not in terms of revenue.

which consequently increases adjustment costs. Threshold-dependent enforcement based on multiple criteria is applied, e.g., in Denmark, Sweden, Germany, Turkey, Russia, Brazil and India (OECD, 2015).

Furthermore, adjustment costs vary in the cross section due to firm-specific heterogeneity. Specifically, as the costs of operationally implementing size management are mostly variable costs (Almunia and Lopez-Rodriguez, 2018), adjustment costs are conditional on the amount by which true, i.e., unmanaged, firm size exceeds the FST. Additional firm-specific heterogeneity in adjustment costs results from internal coordination costs and the quality of a firm's internal information environment (Gallemore and Labro, 2015). Moreover, the level of trust in public institutions affects adjustment costs via social norms. Specifically, a high level of trust in public institutions affects social norms by reducing the willingness of employees to become involved in presumably illegitimate activities (Alm, 2019). Since size management is likely considered illegitimate and because it requires coordination between various employees within a firm, a high level of trust in public institutions increases the adjustment costs of size management.

Information costs are conditional primarily on the amount of information that has to be taken into account by firms to be able to consider all the relevant aspects of an enforcement regime, specifically the segmentation of taxpayers and the audit selection process. Hence, information costs are conditional on the complexity of the threshold-dependent enforcement regime. Imperfect information resulting from information costs can prevent taxpayers from optimal behavior, a phenomenon referred to as inattention in the prior literature (Bosch et al., 2019; Kleven and Waseem, 2013; Kosonen and Matikka, 2019; Søggaard, 2019).⁶ For instance, according to prior research, taxpayers seem to have systematic misperceptions of their average and marginal tax rate, leading to suboptimal tax decisions. This scenario applies to individuals (Brown, 1969; Fujii and Hawley, 1988) as well as to firms (Graham et al., 2017). Furthermore, there is overwhelming evidence that taxpayers tend to subjectively overestimate low probabilities in tax settings, such as the probability of being audited (Alm, 2019).

3 Institutional setting

3.1 Overview

The tax administration in Germany is decentralized to the level of the 16 states. Nonetheless, most taxes are shared between the federal government and the state governments (e.g., personal income tax (PIT), CIT and VAT). Operational tax collection and tax enforcement are conducted by local tax offices, mostly on the level of Germany's approximately 400 districts, and are under supervision by the states' ministries of finance. Comparability of tax enforcement across states is ensured by

⁶ Some literature also uses the term "salience" to describe how tax-relevant information is noticed and acted upon by taxpayers (Hoopes et al., 2015).

federal courts and by binding administrative regulations issued by the Federal Ministry of Finance. However, states may differ particularly in the resources that are available for audits.

3.2 Firm size thresholds

Germany aims to increase the resource efficiency of its tax audits by segmenting firms and by applying different levels of audit intensity to each segment. To this end, firm size is the most relevant segmentation criterion. Specifically, firms are segmented into four size classes (VS-, S-, M- and L-class) based on FSTs that refer to individual legal entities.⁷

FSTs in Germany vary between industries. Specifically, the German tax administration differentiates between four main audit industry clusters (AICs): trading, manufacturing, freelancing and services.⁸ Every three years, i.e., at the beginning of each segmentation cycle, firms that belong to one of these AICs are assigned to a specific size class if their size exceeds either the respective FST for profit or for revenue (or both).⁹

For each segmentation, the tax administration uses information on profit from CIT returns or PIT returns and information on revenue from VAT returns to assign firms to one of the size classes. For the segmentation cycle starting in t the profit and revenue information commonly derive from tax returns for the year $t - 3$ or the year $t - 2$. However, firms cannot know which year's tax return will be used for segmentation. Consequently, firms that intend to engage in size management need to ensure that they do not exceed the respective FST for profit and for revenue in both $t - 3$ and $t - 2$. Furthermore, FSTs are marginally adjusted prior to each segmentation cycle. Although the adjusted FSTs of each segmentation cycle are made publicly available online on the website of the Federal Ministry of Finance and in the Federal Gazette shortly before the segmentation, firms in $t - 3$ and $t - 2$ do not know the exact FSTs that will be applied in the next segmentation cycle starting in t .

Table 1 reports the FSTs between the VS-class and the S-class (VSS-FST), the S-class and the M-class (SM-FST) and the M-class and the L-class (ML-FST) for the main AICs applied for the segmentation cycles starting in 2004 (Panel A), in 2007 (Panel B) and in 2010 (Panel C).

All FSTs invariably increase over time in terms of both profit and revenue. As an example, the VSS-FST for trading firms in 2004 for profit (revenue) was 30,000 (145,000) euros, the SM-FST was 47,000 (760,000) euros, and the ML-FST was 244,000 (6,250,000) euros. By 2010, the VSS-FST increased to 34,000 (160,000) euros, the SM-FST to 53,000 (840,000) euros and the ML-FST to 265,000 (6,900,000) euros.

⁷ See Paragraph 3 of the German Tax Audit Regulation (Betriebsprüfungsordnung). Firm groups are subject to a separate audit target selection scheme that does not rely on FSTs.

⁸ In addition to these four AICs, there are some specific, less-relevant AICs, e.g., financial institutions, insurers, and agricultural and forestry firms. These are not considered here.

⁹ See Paragraph 32(4) of the German Tax Audit Regulation (Betriebsprüfungsordnung).

Table 1 Firm size thresholds 2004, 2007 and 2010

AIC	Criterion	FST (in EUR)			Panel B: Year 2007			Panel C: Year 2010		
		Panel A: Year 2004			VSS-FST	SM-FST	ML-FST	VSS-FST	SM-FST	ML-FST
		VSS-FST	SM-FST	ML-FST						
Trading	Profit	30,000	47,000	244,000	32,000	50,000	250,000	34,000	53,000	265,000
	Revenue	145,000	760,000	6,250,000	155,000	800,000	6,500,000	160,000	840,000	6,900,000
Manufacturing	Profit	30,000	47,000	215,000	32,000	50,000	220,000	34,000	53,000	235,000
	Revenue	145,000	430,000	3,500,000	155,000	450,000	3,700,000	160,000	480,000	4,000,000
Freelancing	Profit	30,000	111,000	485,000	32,000	115,000	500,000	34,000	123,000	540,000
	Revenue	145,000	700,000	3,700,000	155,000	735,000	3,900,000	160,000	790,000	4,300,000
Services	Profit	30,000	51,000	265,000	32,000	55,000	280,000	34,000	59,000	305,000
	Revenue	145,000	630,000	4,700,000	155,000	660,000	4,900,000	160,000	710,000	5,300,000

This table reports the VSS-FST, SM-FST and ML-FST applied for the segmentation cycle starting in 2004 (Panel A), in 2007 (Panel B) and in 2010 (Panel C) for the four main AICs: trading, manufacturing, freelancing and services in terms of profit and revenue. Data: German Federal Ministry of Finance (2003, 2006, 2009)

Table 2 Changes in firm size thresholds 2004–2007 and 2007–2010

Audit Group	Criterion	Changes in FSTs (in EUR)					
		Panel A: 2004 to 2007			Panel B: 2007 to 2010		
		VSS-FST	SM-FST	ML-FST	VSS-FST	SM-FST	ML-FST
Trading	Profit	2000 (6.7%)	3000 (6.4%)	6000 (2.5%)	2000 (6.3%)	3000 (6.0%)	15,000 (6.0%)
	Revenue	10,000 (6.9%)	40,000 (5.3%)	250,000 (4.0%)	5000 (3.2%)	40,000 (5.0%)	400,000 (6.2%)
Manufacturing	Profit	2000 (6.7%)	3000 (6.4%)	5,000 (2.3%)	2000 (6.3%)	3000 (6.0%)	15,000 (6.8%)
	Revenue	10,000 (6.9%)	20,000 (4.7%)	200,000 (5.7%)	5000 (3.2%)	30,000 (6.7%)	300,000 (8.1%)
Freelancing	Profit	2000 (6.7%)	4000 (3.6%)	15,000 (3.1%)	2000 (6.3%)	8000 (7.0%)	40,000 (8.0%)
	Revenue	10,000 (6.9%)	35,000 (5.0%)	200,000 (5.4%)	5000 (3.2%)	55,000 (7.5%)	400,000 (10.3%)
Services	Profit	2000 (6.7%)	4000 (7.8%)	15,000 (5.7%)	2000 (6.3%)	4000 (7.3%)	25,000 (8.9%)
	Revenue	10,000 (6.9%)	30,000 (4.8%)	200,000 (4.3%)	5000 (3.2%)	50,000 (7.6%)	400,000 (8.2%)

This table reports the euro and percentage changes (in parentheses) in VSS-FSTs, SM-FSTs and ML-FSTs from 2004 to 2007 (Panel A) and from 2007 to 2010 (Panel B) for the four main AICs: trading, manufacturing, freelancing and services in terms of profit and revenue using the information reported in Table 1

Table 2 reports the euro and percentage changes (in parentheses) in FSTs from 2004 to 2007 (Panel A) and from 2007 to 2010 (Panel B) for the main AICs using the information reported in Table 1.

Across all AICs, neither the percentage nor the euro adjustments of the FSTs are consistent over time. For instance, the VSS-FST for trading firms increased by 2,000 (10,000) euros, the SM-FST by 3,000 (40,000) euros and the ML-FST by 6,000 (250,000) euros for profit (revenue) from 2004 to 2007. From 2007 to 2010, the VSS-FST increased by 2,000 (5,000) euros, the SM-FST by 3,000 (40,000) euros and the ML-FST by 15,000 (400,000) euros. In relative terms, the increases range from 2.5% to 6.9%. Consequently, it is not possible for firms to exactly predict the FSTs that will be applied in the next segmentation cycle. However, firms are aware of FSTs applied for the current segmentation cycle, and FSTs have historically never decreased.¹⁰ Consequently, firms using a conservative approach will rationally

¹⁰ Once published by the Federal Ministry of Finance, the FSTs are also covered by professional media. Hence, it is rather easy for firms to become aware of the FSTs and to access information on the FSTs applied for the current segmentation cycle.

manage their size to the FSTs last made publicly available, i.e., FSTs applied for the current segmentation cycle.

3.3 Audit probability

A firm's size class strongly affects its audit probability due to the specific design of the audit target selection process, which relies on 1) risk-dependent selection, 2) random selection and 3) time-dependent selection (Harle and Olles, 2017; Wenzig, 2014). First, under risk-dependent selection, firms are selected for audit based on firm-specific risk factors identified from entries in tax returns. These risk factors include, e.g., foreign business activities, loss carry-forwards and deviations from industry averages. Second, under random selection, firms are drawn randomly and independently of firm-specific characteristics. Specifically, within each size class, a number of firms are drawn randomly to reduce the predictability of audits. Finally, and most important, under time-dependent selection, firms are selected regardless of their firm-specific characteristics but only according to binding target intervals at which firms in each size class must be audited. These target intervals differ across size classes and are three to four years for L-class firms, 8.5 to 10.5 years for M-class firms and 14.4 to 20 years for S-class firms. For VS-class firms, no target interval is set (Bavarian General Accounting Office, 2013; Kaligin, 2014).

Despite a slight increase in the application of risk-dependent selection since the introduction of automated risk management systems in recent years, time-dependent selection remains the most important component of the target selection process in Germany (German Bundestag, 2021; Klein and Rüsken, 2020). Because time-dependent selection depends exclusively on a firm's size class, size class is the major determinant of audit probability. Furthermore, as target intervals differ across the size classes, audit probability changes discontinuously at FSTs. Coherently, eight out of nine tax consulting professionals consider a firm's size class as the major determinant of audit probability in Germany.¹¹

Note that the amount by which a respective FST is exceeded is irrelevant for size class segmentation and that individual auditors have little discretion in selecting firms because audit schedules are established at the level of local tax offices according to the target selection process described above. Nevertheless, due to risk-dependent selection, and the fact that firm size is presumably positively correlated with some risk factors, audit probability positively correlates with firm size within size classes. This may attenuate the discontinuities in audit probability at FSTs to a certain degree. However, because time-dependent selection represents by far the most important component of the target selection process and because target intervals vary substantially across size classes, it is very unlikely that risk-dependent selection would completely eliminate the jumps in audit probability.

Historical audit rates conditional on size class are made publicly available on an annual basis by the Federal Ministry of Finance. As audit rates remain virtually

¹¹ See Henselmann and Haller (2017) for survey results and Panek (2018), Strangmeier (2000) and Wenzig (2014) for a discussion of the literature.

unchanged over time, they provide a reliable estimate of firms' audit probabilities. According to several rulings of the German Federal Finance Court, the differences in audit rates across size classes do not violate the principle of equality of the German constitution because the tax administration is allowed to segment taxpayers for an effective use of its limited resources.¹²

3.4 Audit quality

Size class also affects audit quality. Specifically, administrative regulation dictates that for L-class firms, audits must be consecutive, i.e., once a firm is audited, it must cover all years that were not covered by the previous audit of that firm, whereas for M-class, S-class and VS-class firms, the audit period must not exceed three calendar years. Moreover, more-experienced and better-trained auditors are generally assigned to larger cases (Bavarian General Accounting Office, 2013). Additionally, the size and specialization of audit teams increase with the audited firm's size class. Furthermore, the Federal Central Tax Office regularly assigns additional federal auditors to audits of mainly L-class firms.

3.5 Audit outcomes

Table 3 reports historical audit rates, audit periods and additional tax revenues generated from audits (consisting of additional tax claims, interest payments and penalty fees) per size class for the years 2004 (Panel A), 2007 (Panel B) and 2010 (Panel C).

The majority of German firms are assigned to the VS-class, which is expected. For instance, in 2010, 74.6% of firms were assigned to the VS-class, 13.9% to the S-class, 9.3% to the M-class and 2.2% to the L-class. Furthermore, it can be seen that audit rates change strongly at FSTs. In 2010, 1.0% of firms in the VS-class, 3.5% of firms in the S-class, 6.9% of firms in the M-class and 21.1% of firms in the L-class were audited. On average, an audit covered 2.9 calendar years in the VS-class and the S-class, 3.0 years in the M-class and 3.3 calendar years in the L-class.

Consequently, 70.8% of the additional tax revenue of 16.8 billion euros was derived from audits of L-class firms in 2010. This corresponds to 293,813 euros per audited firm. However, with 15,013 (16,878) [23,502] euros, the average additional tax revenue per audited firm was economically significant for the VS-class (S-class) [M-class] as well.¹³

3.6 Benefits of size management

As discussed in Sect. 2.2.1, firms engage in size management around FSTs only if the benefits of size management, i.e., the difference in expected firm audit costs just

¹² For instance, see German Federal Court of Finance (1988).

¹³ A recent survey among German firm managers shows that approximately 75% of all audits result in additional tax revenue (PricewaterhouseCoopers, 2019).

above and just below the FST, exceed optimization costs. To provide some indication of the extent of the benefits of size management, we conduct a simple back-of-the-envelope calculation.

First, we assume that conditional firm audit costs do not change strongly at FSTs, i.e., between size classes. This is a simplification, as size class particularly affects audit quality (see Sect. 3.4). Under this assumption, the decrease in expected firm audit costs that is caused by size management results merely from the discontinuous changes in audit probability at FSTs. We use the average additional tax revenue per audited firm in 2010 from Table 3 as a proxy for conditional firm audit costs, specifically, additional tax claims, interest payments and penalty fees, for the average firm in each size class.¹⁴ We further make the simplifying assumption that the profit of the average firm in each size class corresponds to the midpoint of that size class. The profits of the smallest and the largest firm in each size class are defined by the FSTs for trading firms for the segmentation cycle starting in 2010 from Table 1. For the VS-class, we assume that the smallest firm in that size class makes a profit of zero, and for the L-class, we assume that the largest firm makes a profit of ten million euros.¹⁵

We divide the conditional firm audit costs for the average firm in each size class by the profit of the average firm to obtain the ratio of conditional firm audit costs to profits for the average firm in each size class. To obtain the ratio of conditional firm audit costs to profits at FSTs, we calculate the mean of the ratio of conditional firm audit costs to profits of the average firm in the size class to the left and to the right of the respective FST. To finally derive the ratio of expected firm audit costs to profits at FSTs, we multiply the conditional firm audit costs to profits at FSTs by the audit rates, i.e., a proxy for audit probabilities, in the size class to the left and to the right of the respective FST. As audits usually cover more than one calendar year, we multiply audit rates by the average audit period in each size class in 2010 from Table 3 to obtain proxies for the probability that the tax return for a single year will be audited. Correspondingly, we divide conditional firm audit costs and expected firm audit costs by the average audit period to obtain conditional and expected audit costs per year.

To account for the possibility that audit probability is correlated with firm size within size classes, we assume that audit rates for the smallest (largest) firms in every size class correspond to 90% (130%) of the average audit rate in that size class in 2010 from Table 3.¹⁶

¹⁴ Note that compliance costs are not included in our estimate. However, the compliance costs of audits might be substantial in the German setting. Specifically, if we multiply the number of hours usually charged by tax consultants to accompany a tax audit, i.e., 35 hours (Meyer, 1988), by the standard hourly fee of 140 euros according to Paragraph 29 of the German Tax Consultant Fees Regulation (Steuerberatervergütungsverordnung), this amounts to 4,900 euros per audit.

¹⁵ In 2010, the vast majority of firms in Germany had a revenue below 50 millions of euros, and the return on sales was, on average, approximately 5% (Kreditanstalt für Wiederaufbau, 2012).

¹⁶ This assumption is based on the variation in audit rates for the smallest and largest firms in the L-class in Rhineland-Palatinate in 2013 (Regional Tax Authority of Rhineland-Palatinate, 2016), which is the only information on audit rates within size classes available. Specifically, no fine-grained information is available for other size classes or other German states.

Table 3 Audit outcomes 2004, 2007 and 2010

Size class	Firms		Audit rate	Audit period	Additional tax revenue		
	N	%			%	years	EUR total (in millions)
<i>Panel A: Year 2004</i>							
VS	5,252,015	71.6	1.8		875	6.6	12,798
S	1,111,628	15.2			617	4.6	
M	795,073	10.8	7.8		1,264	9.5	20,263
L	172,184	2.3	22.9		10,547	79.3	266,978
Total	7,330,900	100.0	3.0		13,303	100.0	60,894
<i>Panel B: Year 2007</i>							
VS	6,284,418	75.2	1.1	2.9	820	4.9	11,564
S	1,140,402	13.7	3.9	3.0	630	3.8	14,083
M	757,810	9.1	7.8	3.0	1,390	8.4	23,532
L	169,843	2.0	22.8	3.5	13,200	79.5	341,421
Total	8,352,473	100.0	2.6		16,040	100.0	77,797
<i>Panel C: Year 2010</i>							
VS	6,391,015	74.6	1.0	2.9	1,000	6.0	15,013
S	1,189,727	13.9	3.5	2.9	700	4.2	16,878
M	799,135	9.3	6.9	3.0	1,300	7.7	23,502
L	191,638	2.2	21.1	3.3	11,900	70.8	293,813
Total	8,571,515	100.0	2.4		16,800	100.0	82,392

This table reports the historical audit rates, audit periods and additional tax revenues generated by audits for the years 2004 (Panel A), 2007 (Panel B) and 2010 (Panel C) for VS-class, S-class, M-class and L-class firms. In 2004, the available data on audit rates do not differentiate between VS-class and S-class firms, and no information on audit periods is available. Data: German Federal Ministry of Finance (2005, 2008, 2011)

Figure 1 shows the ratio of conditional firm audit costs to profits (dot markers) for the average firm in the VS-class, S-class, M-class and L-class and at the VSS-FST, SM-FST and ML-FST (solid horizontal lines) under these assumptions. The short-dashed line represents a trend line of the ratio of conditional firm audit costs to profits based on a third-order polynomial. The dash-dotted line indicates the audit probability. Finally, the solid line shows the ratio of expected firm audit costs to profits at FSTs, i.e., the ratio of conditional firm audit costs to profits multiplied by the audit rate in the respective size classes.

The ratio of conditional firm audit costs to profits is 30.5% (13.4%) [4.9%] {1.7%} for the average firm in the VS-class (S-class) [M-class] {L-class}. Hence, our estimates indicate that the ratio of conditional firm audit costs to profits is decreasing with firm size. This is plausible for two reasons. First, larger firms tend to have more tax expertise and hence likely engage in more sound tax avoidance compared to smaller firms (Chen et al., 2010). As more sound tax avoidance is less likely to be objected to by the tax administration, this leads to a lower ratio of conditional firm audit costs to profits. Second, a decreasing ratio of conditional firm audit costs to

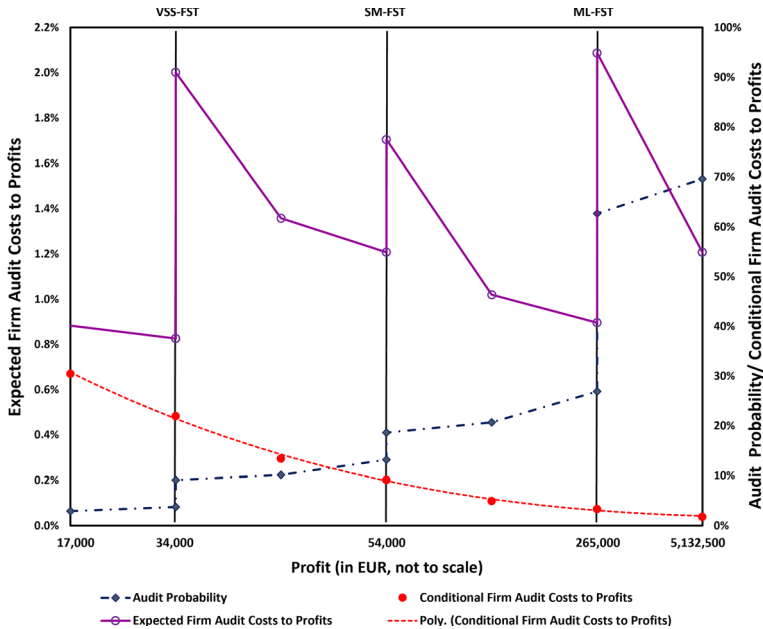


Fig. 1 Benefits of size management at firm size thresholds. *Notes:* This figure graphs the ratio of conditional firm audit costs, i.e., firm audit costs once a firm is audited, to profits (dot markers) for the average firms in the VS-class, S-class, M-class and L-class and for firms at the VSS-FST, SM-FST and ML-FST (vertical solid lines) for the segmentation cycle starting in 2010 obtained from the back-of-the-envelope calculation described in Sect. 3.6. The short-dashed line represents a trend line of the ratio of conditional firm audit costs to profits based on a third-order polynomial. The dash-dotted line indicates audit probabilities in the individual size classes in 2010 using information from Table 3. The solid line shows the ratio of expected firm audit costs to profits, i.e., the ratio of conditional firm audit costs to profits multiplied by the audit probability in the respective size classes. As audits usually cover more than one calendar year, we multiply audit rates by the average audit periods in each size class in 2010 from Table 3 to obtain proxies for the probability that the tax return for a single year will be audited. Correspondingly, we divide conditional firm audit costs and expected firm audit costs by the average audit period to obtain conditional and expected audit costs per year

profits is also consistent with the political cost hypothesis. The political cost hypothesis predicts that larger firms take less aggressive tax positions (Gupta and Newberry, 1997; Zimmerman, 1983). Less aggressive tax positions also imply a lower probability of objections by the tax administration and hence lower conditional firm audit costs relative to profits.

The ratio of conditional firm audit costs to profits is 21.9% at the VSS-FST. Accordingly, the ratio of expected firm audit costs to profits is 0.8% just below the VSS-FST, where audit probability is 3.8%, and 2.0% just above the VSS-FST, where audit probability is 9.1%. Consequently, expected firm audit costs decrease by 1.2% of profits if a firm with profit just above the VSS-FST engages in size management. Analogously, the decrease in the ratio of expected firm audit costs to profits due to size management is 0.5% at the SM-FST and 1.2% at the ML-FST. Hence, the

benefits of size management appear to be substantial in economic terms at all FSTs, and therefore, firms have reason to manage their size at FSTs.

4 Hypothesis development

Despite the substantial benefits of size management at all FSTs, it remains an empirical question whether firms in our setting engage in size management, as no data are available to provide a reliable estimate of optimization costs for German firms. However, the criteria applied for segmentation and the high complexity of the threshold-dependent enforcement regime in Germany are expected to increase optimization costs as described in Sect. 2.2.3. First, firms in Germany have to take into account multiple criteria, i.e., profit and revenue, in their size management, which makes size management considerably more difficult and more time-consuming. Second, profit and revenue are more difficult to manage than profit alone, as revenue cannot be adjusted through additional expenditures at the last minute. Finally, the complexity of the enforcement regime, e.g., four different size classes, regular adjustments of FSTs and industry-specific FSTs, also make FSTs less salient for firms and increase optimization cost via the information costs channel.

Accordingly, we state H1 as follows in the alternative form:

Hypothesis 1 Threshold-dependent tax enforcement is associated with size management.

As shown in Fig. 1, the benefits of size management, i.e., the decreases in expected firm audit costs, vary between FSTs. However, the variation is not substantial. Accordingly, we state H2 as follows in the alternative form:

Hypothesis 2 The extent of size management, i.e., the number of firms engaged in size management relative to the total number of firms around that FST, varies between size classes.

As discussed in Sect. 2.2.2, conditional firm audit costs vary between industries. For instance, under third-party reporting, the traceability of transactions is presumably larger in industries with a major share of business customers compared to industries with a major share of individual customers. Hence, incentives to engage in size management vary between AICs. Accordingly, we state H3 as follows in the alternative form:

Hypothesis 3 The extent of size management varies between AICs.

5 Empirical strategy

To test our hypotheses, we exploit the fact that size management creates a discontinuity around the FST in an otherwise relatively smooth firm size distribution. More specifically, size management creates a missing mass (smaller number of firms than any continuous distribution would predict) above the FST and an excess mass (larger number of firms than any continuous distribution would predict) below it. Due to variable adjustment costs, the missing mass is expected to derive from a limited area above the FST. Furthermore, also due to variable adjustment costs, the excess mass is expected to be located in a limited area below the FST.¹⁷

To test H1 and H2, we fit a polynomial to the distribution of *SIZE*, which denotes profit (*EBT*) and revenue (*REV*), i.e., the two size variables on which FSTs are based in our setting. Technically, both *EBT* and *REV* are standardized by dividing all values of *SIZE* by the FST last made publicly available for the respective AIC, i.e., the standardized variables take a value one if a firm exactly meets the FST.¹⁸

We adapt techniques from prior bunching literature (Chetty et al., 2011; Kleven and Waseem, 2013; Saez, 2010).¹⁹ Specifically, we divide *SIZE* into equal-sized bins and fit a fifth-order polynomial using the midpoint of each bin as data points. We estimate a regression of the following form:

$$F_j = \sum_{i=0}^5 \beta_i \cdot (x_j)^i + \sum_{k=x_{lb}}^{x_{ub}} \gamma_k \cdot \mathbb{1}(x_j = k) + \epsilon_j, \quad (1)$$

where F_j is the percentage of firms in bin j (i.e., relative to the total number of firms in all bins), x_j is the *SIZE* midpoint of bin j and the γ_k 's are intercept shifters, i.e., coefficients for each of the bins in the bunching interval, i.e., the area where bunching is expected. The indicator function $\mathbb{1}(x_j = k)$ takes the value one for each of the bins in the bunching interval with x_{lb} and x_{ub} being the lower and upper bounds of the bunching interval, respectively. Consistent with Bernard et al. (2018), we choose the bin width as 2% of the FST. This bin width is large enough for the distributions of *SIZE* to be relatively smooth (in the absence of size management) but presumably small enough for firms to manage size by the amount corresponding to the bin width at a reasonably low cost.²⁰ Following Almunia and Lopez-Rodriguez (2018) and Bernard et al. (2018), we focus on firms in the interval between 50 and 150% of

¹⁷ The excess mass is not expected to form a single spike at the FST, as firms are unable to manage size to exactly match the FST, e.g., due to the indivisibility of transactions (Almunia and Lopez-Rodriguez, 2018).

¹⁸ Recall that at the time firms have to engage in size management, firms do not know the exact FSTs that will be applied in the next segmentation cycle, but firms are aware of FSTs applied for the current segmentation cycle and know that FSTs have never been adjusted downward in the past. Consequently, it is reasonable to assume that using a conservative approach, firms will manage their size to the FSTs applied for the current segmentation cycle (see Sect. 2.1). However, we repeat our analyses at different placebo FSTs to ensure that our results are not driven by the selection of FSTs (Sect. 8.6).

¹⁹ We apply alternative bunching tests in Online Appendix A.4 to corroborate our results.

²⁰ We apply two different specifications of the bin width, i.e., 0.5–1%, in Online Appendix A.3 to ensure that our results are not driven by the model specification.

each FST to obtain precise estimates. We set the lower bound of the bunching interval as three bins to the left and the upper bound as three bins to the right of the FST.

H1 predicts size management to occur around the FSTs. H1 is confirmed if any of the γ_k s to the left ($\gamma_{0.95}$, $\gamma_{0.97}$, $\gamma_{0.99}$) are positive and significant, indicating an excess mass below the respective FST. Furthermore, the coefficients are expected to decrease in absolute values with increasing distance to the FST due to increasing optimization costs.²¹

H2 predicts that the extent of size management varies between size classes. H2 is confirmed if the γ_k s to the left differ significantly across individual FSTs.

H3 predicts that the extent of size management varies between AICs. To test H3, we estimate Equation 1 separately for each of the four main AICs for both *EBT* and *REV*. H3 is confirmed if the γ_k s to the left differ significantly across individual AICs. To control for differences between industries within AICs, we also estimate Equation 1 separately for every individual industry as defined by the 2-digit NACE code.²² Again, H3 is confirmed if the γ_k s to the left differ significantly across individual industries.

6 Data

6.1 Sample selection

We obtain administrative microlevel tax return data for 2010 on the entire population of German firms from the Research Data Center (RDC) of the Federal Statistical Office and the Statistical Offices of the Federal States.²³ All data are taken from the firms' submitted tax returns, i.e., the data are prior to changes induced by audits. Specifically, we obtain data on the CIT of corporations, PIT of partners in partnerships and local business tax (LBT) of corporations, partnerships and sole proprietors.²⁴ We also obtain data on both annual VAT returns and VAT prefilling returns (prefillings usually occur monthly or quarterly). Table 4 shows the sample selection process.

The data originally include 2,756,463 firms with information on both *REV* and *EBT*.²⁵ We first exclude 36,571 (1.33%) firms that belong to a fiscal unity group for either CIT, LBT or VAT, as the FSTs refer to individual legal entities, while the

²¹ Negative and significant γ_k s to the right ($\gamma_{1.01}$, $\gamma_{1.03}$, $\gamma_{1.05}$) indicating a missing mass are not required to confirm H1 as the missing mass might be dispersed across a larger area.

²² We show that for most industries the sample size is large enough to keep the probability of making a type II error below 1% in Online Appendix A.2.

²³ We repeat our analyses with data for 2004 and 2007 to ensure that our results are not only prevalent in 2010 (Sect. 8.2).

²⁴ Partnerships and sole proprietors in certain industries, such as legal consulting and agricultural or forestry firms do not pay local business tax and are thus not included in the data.

²⁵ The raw data also include 5,183,225 firms with a missing entry for *REV* and/or *EBT* in 2010. These firms are excluded altogether.

available data contain information only on profit and revenue aggregated at the fiscal unity level.

The data in principle contain information about the exact AICs to which a firm is allocated by the tax administration (i.e., trading, manufacturing, freelancing or services). However, for some of the firms, this information is missing. If this is the case, we use 5-digit NACE codes and information on legal form and LBT liability to allocate firms to the correct AICs. We ultimately exclude 6,934 firms (0.25%) that cannot be allocated to a unique AIC with the available information and 32,403 (1.18%) firms that do not belong to one of the four main AICs. Finally, we exclude all industries as defined by the 2-digit NACE code with fewer than 50 observations in the interval between 50 and 150% of each FST for *EBT* and *REV* so that, on average, we have at least one observation for each of the 50 bins used in the regression for any industry-specific analysis. This process excludes 678 (0.02%) firms. Our final sample contains 2,679,877 (97.22%) firms.²⁶

If available, we use the reported profit of either the CIT return or the PIT return as our *EBT* variable, which is also the variable definition used by the tax administration. If neither of these variables is available, we use the profit reported on the LBT return, which is closely associated with the profit of the CIT or the PIT returns. As our *REV* variable, we use revenues reported on annual VAT returns, which is again the variable definition used by the tax administration. If this variable is not available, we use firm-level cumulated revenues as reported on all 2010 prefilling VAT returns.

6.2 Descriptive statistics

Descriptive statistics of raw, i.e., nonstandardized, *EBT* (*rawEBT*) and raw *REV* (*rawREV*) are reported in Table 5. We report nonstandardized values of *SIZE* here to allow a better understanding of the data. We also report in Table 5 the exact tax returns that are used to collect *rawEBT* and *rawREV*.

The average firm reports a *rawEBT* of 60,240 euros (median: 16,224 euros). *rawEBT* is based on CIT data in 23.26% of cases, PIT data in 18.56% of cases and LBT data in 58.19% of cases. The average *rawREV* is 941,742 euros (median: 101,390 euros). *rawREV* is based on VAT returns in 99.73% of cases and VAT prefilling returns in 0.27% of cases.

We further provide a naive graphical assessment of the distributions of *EBT* and *REV*. Figures 2 and 3 show the firm size distribution of *EBT* and *REV*, respectively, around the VSS-FSTs, SM-FSTs and ML-FSTs for the segmentation cycle starting in 2010 (solid vertical line) for the overall population of firms and separately for each AIC. The bin width is set to 2% of the FSTs. The bunching interval is set to three bins to the right and three bins to the left (dashed vertical lines).

²⁶ As a robustness test, we restrict our sample to firms not exceeding the respective other FST to reduce noise in our analyses (Sect. 8.3). Furthermore, we restrict our sample to loss firms because financially constrained firms likely have larger incentives to engage in size management (Sect. 8.4). We also repeat our analyses for individual states and for individual districts to control for geographic heterogeneity in tax enforcement across Germany (Sect. 8.5).

Table 4 Sample selection

Initial sample (firms with information on both <i>REV</i> and <i>EBT</i>)	2,756,463	100.0%
Firms belonging to a fiscal unity group	−36,571	1.33%
Firms without AIC information	−6,934	0.25%
Firms not belonging to one of the main AICs	−32,403	1.18%
Firms from industries with less than 50 observations in the interval [0.5, 1.5] around each FST for <i>EBT</i> and <i>REV</i>	− 678	0.02%
Final sample	2,679,877	97.22%

The distributions of both *EBT* and *REV* are relatively smooth and decrease in firm size around all FSTs. The distributions also become more convex for smaller FSTs. There are no notable discontinuities at any of the FSTs, neither for the full sample of firms nor when considering the four AICs separately. In Fig. 2, we note that *EBT* has some visible spikes in the distributions (while *REV* does not). However, these spikes in *EBT* do not appear to be associated with size management, as they seem to be distributed at random.

7 Results

Tables 6 and 7 report the regression results from estimating Equation 1 for *EBT* and *REV*, respectively. Panel A presents findings for H1 and H2, i.e., the results for the full sample of firms at the VSS-FSTs, SM-FSTs and the ML-FSTs for the segmentation cycle starting in 2010. Panel B presents AIC-specific findings for H3, i.e., subsample results per AIC at the VSS-FSTs, SM-FSTs and ML-FSTs. We report the coefficients for three bins to the left of the FSTs ($\gamma_{0.95}$, $\gamma_{0.97}$, $\gamma_{0.99}$) and three bins to the right ($\gamma_{1.01}$, $\gamma_{1.03}$, $\gamma_{1.05}$) in Panel A and the coefficient of the first bin to the left ($\gamma_{0.99}$) and the first bin to the right of the FSTs ($\gamma_{1.01}$) in Panel B.

All coefficients but one are economically small and statistically nonsignificant for the full sample of firms reported in Panel A of Tables 6 and 7. Hence, our data do not support H1, i.e., we do not find evidence of size management around FSTs. Consequently, the first implication of our results is that for German firms, optimization costs exceed the benefits of size management. Furthermore, as the coefficients are nonsignificant across all FSTs, the data also do not support H2, i.e., our results imply that optimization costs exceed the benefits in all size classes despite heterogeneity in benefits between those size classes. Along the same lines, three out of 48 coefficients per AIC reported in Panel B of the tables are economically small and statistically nonsignificant at the 10% level, which implies that optimization costs exceed benefits in all AICs. Hence, optimization costs appear to be considerably large.

Figures 4 and 5 present the industry-specific findings for H3, i.e., subsample results per industry at the VSS-FSTs, SM-FSTs and ML-FSTs for the segmentation cycle starting in 2010. Note that under the null of an absence of size management, coefficients are asymptotically normally distributed around zero. Consequently, t

Table 5 Descriptive statistics

SIZE	N	%	Mean	SD	Q1	Median	Q3
<i>rawEBT</i>	2,679,877	100.00	60,240	1,574,183	413	16,224	42,649
<i>from CIT returns</i>	623,221	23.26	62,218	1,564,167	-3,535	4,161	30,652
<i>from PIT returns</i>	497,327	18.56	144,568	1,901,604	161	19,744	82,174
<i>from LBT returns</i>	1,559,329	58.19	32,555	1,457,607	2,929	20,213	40,000
<i>raw REV</i>	2,679,877	100.00	941,742	1.93e+07	30,553	101,390	337,355
<i>from VAT returns</i>	2,672,639	99.73	941,361	1.93e+07	30,492	101,311	337,037
<i>from VAT prefilling returns</i>	7,238	0.27	1,082,396	6,992,035	49,207	132,396	486,533

This table presents descriptive statistics for *rawEBT* and *rawREV* and shows how *rawEBT* and *rawREV* are composed of the distinct profit and revenue variables available in the data. Further, it shows the descriptives of the variables used to construct *rawEBT* or *rawREV*

values are asymptotically standard normally distributed, and p values are asymptotically uniformly distributed between zero and one.

In Panel A of the tables, we plot histograms and kernel estimates of density (solid line) for the regression coefficients to the left ($\gamma_{0,99}$) for each of more than 70 industries in our sample. In Panel B of the tables, we plot kernel density estimates (solid line) for the respective t values and compare them to a standard normal density distribution (dashed line) to determine how the empirical distributions of t values fit the theoretical distribution of t values under the null. Finally, in Panel C of the tables, we plot the empirical cumulative distribution functions (ECDFs) for the respective p values. If the p values are distributed uniformly, the ECDF (short-dashed line) follows the line of equality (solid line diagonal).

In Panel A and Panel B of the tables, the vertical axis presents the (empirical) density. In Panel C of the tables, the vertical axis presents the (empirical) cumulative probability. The horizontal axis shows the coefficients, t values and p values.

Panel A of Figs. 4 and 5 shows a symmetric density distribution for the regression coefficients centered around zero for all FSTs and for both *EBT* and *REV*. Furthermore, the empirical density distributions for t values in Panel B fit well with the theoretical density distribution under the null. Additionally, the ECDFs for p values in Panel C follow the line of equality for all FSTs and for both *EBT* and *REV*. Hence, consistent with the AIC-specific findings for H3, the results imply that optimization costs exceed the benefits of size management even when controlling for industry-specific heterogeneity in conditional firm audit costs. Overall, our data do not support H3. Furthermore, as the density distributions of the coefficients are centered around zero, we find an indication that our results are not caused by particularly large standard errors but that coefficients are, in fact, very close to zero. We find virtually similar results for the first coefficients to the right of the FSTs ($\gamma_{1,01}$) (not graphed).

Considered jointly, our data do not support a rejection of the null of an absence of size management at FSTs. This is true for both *EBT* and *REV*. Our results further suggest that optimization costs exceed the benefits of size management even when controlling for size class-specific and industry-specific heterogeneity in

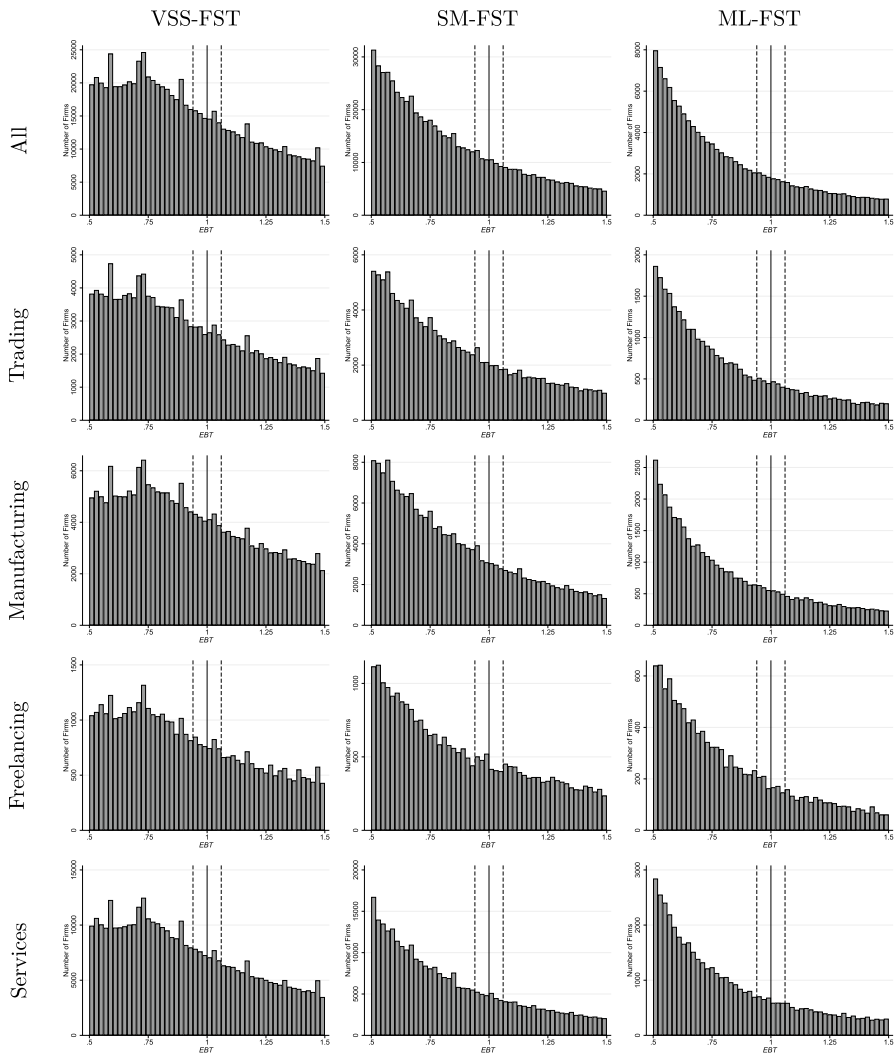


Fig. 2 Distribution of *EBT*. *Notes:* This figure graphs the distribution of *EBT* for the full sample of firms (all) and per AIC (trading, manufacturing, freelancing, services). We focus on firms in the interval [0.5, 1.5] around the VSS-FST, SM-FST and ML-FST for the segmentation cycle starting in 2010, such that the FSTs (solid line) are in the center of the graphs. Bin width is 2% of the FSTs. The bunching interval is set to three bins to the right and three bins to the left (dashed vertical lines)

conditional firm audit costs. Our results correspond to the results found by Tenant and Tracey (2019) for firms in Jamaica and are in contrast to the results found by Almunia and Lopez-Rodriguez (2018) for Spanish firms.

Accordingly, we argue that a pattern seems to be emerging from this relatively new field of research on how the specific design of threshold-dependent policies, i.e.,

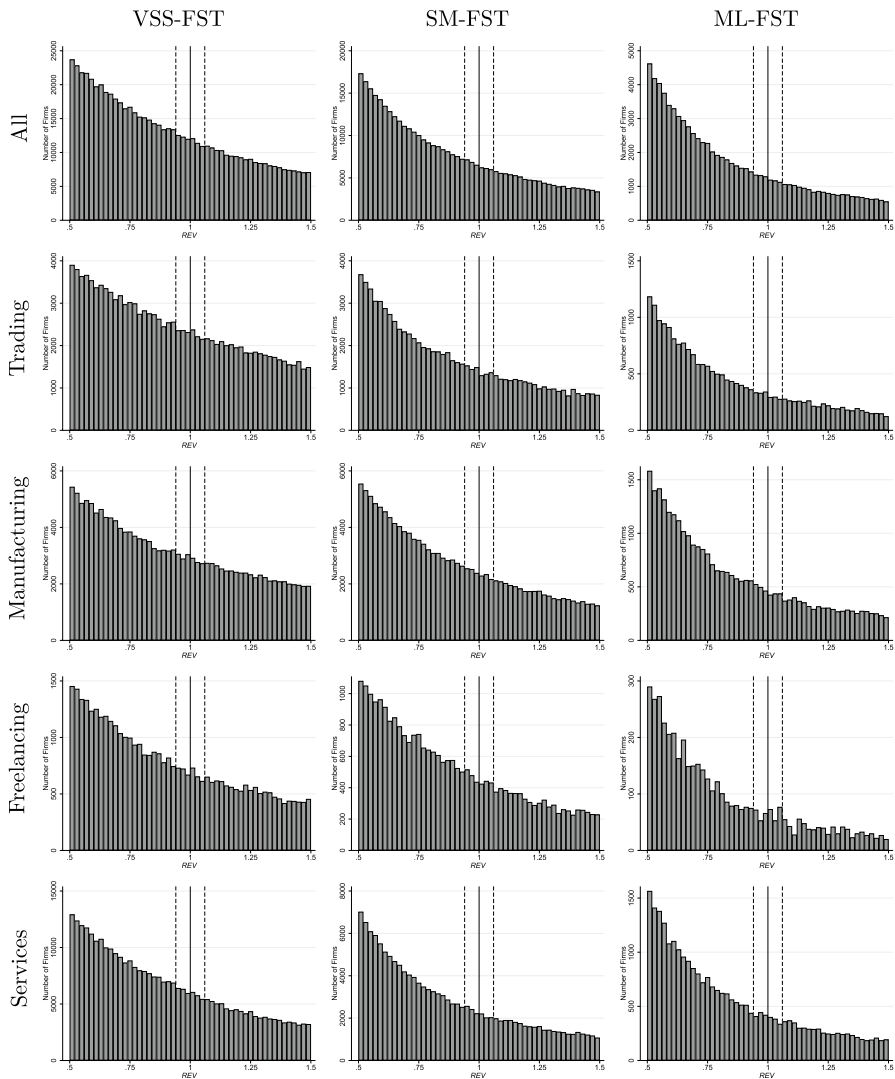


Fig. 3 Distribution of *REV*. *Notes:* This figure graphs the distribution of *REV* for the full sample of firms (all) and per AIC (trading, manufacturing, freelancing, services). We focus on firms in the interval [0.5, 1.5] around the VSS-FST, SM-FST and ML-FST for the segmentation cycle starting in 2010, such that the FSTs (solid line) are in the center of the graphs. Bin width is 2% of the FSTs. The bunching interval is set to three bins to the right and three bins to the left (dashed vertical lines)

the criteria applied for segmentation and the complexity of the threshold-dependent enforcement regime, can inhibit size management. Specifically, we note that Germany and Spain are relatively similar in important drivers of optimization costs because they

Table 6 Results for H1, H2, H3 (per AIC): *EBT*

Panel A: Full sample						
FST	Left			Right		
	$\gamma_{0.95}$	$\gamma_{0.97}$	$\gamma_{0.99}$	$\gamma_{1.01}$	$\gamma_{1.03}$	$\gamma_{1.05}$
VSS	-0.1080 (-0.59)	-0.0915 (-0.50)	-0.1020 (-0.56)	-0.0960 (-0.52)	0.1490 (0.81)	-0.0145 (-0.08)
SM	0.1050 (1.26)	-0.0655 (-0.78)	-0.0264 (-0.31)	0.0246 (0.29)	-0.0158 (-0.19)	-0.0456 (-0.54)
ML	0.0116 (-0.26)	-0.0069 (-0.15)	-0.0138 (-0.31)	-0.0089 (-0.20)	0.0240 (0.53)	0.0015 (0.03)
Panel B: per AIC						
AIC	VSS-FST		SM-FST		ML-FST	
	$\gamma_{0.99}$	$\gamma_{1.01}$	$\gamma_{0.99}$	$\gamma_{1.01}$	$\gamma_{0.99}$	$\gamma_{1.01}$
Trading	-0.1160 (-0.63)	-0.0712 (-0.39)	-0.0021 (-0.02)	-0.0557 (-0.49)	-0.0268 (-0.38)	0.0945 (1.36)
Manufacturing	-0.0931 (-0.52)	-0.0479 (-0.27)	-0.0682 (-0.65)	-0.0437 (-0.41)	-0.0291 (-0.31)	0.0206 (0.22)
Freelancing	-0.0879 (-0.51)	-0.1080 (-0.62)	0.2390** (2.43)	-0.116 (-1.18)	-0.1380 (-0.98)	-0.0342 (-0.24)
Services	-0.1030 (-0.53)	-0.1300 (-0.67)	-0.0338 (-0.26)	0.1130 (0.86)	0.0416 (0.50)	-0.1000 (-1.20)

This table reports the regression coefficients from estimating Equation 1 for *EBT*. Panel A presents findings for H1 and H2, i.e., results for the full sample of firms at the VSS-FST, SM-FST and the ML-FST for the segmentation cycle starting in 2010. Panel B presents the findings for H3, i.e., subsample results at the VSS-FST, SM-FST and ML-FST per AIC. We report the coefficients for three bins to the left of the FST ($\gamma_{0.95}$, $\gamma_{0.97}$, $\gamma_{0.99}$) and all three bins to the right ($\gamma_{1.01}$, $\gamma_{1.03}$, $\gamma_{1.05}$) in Panel A but only the coefficient of the first bin to the left ($\gamma_{0.99}$) and the first bin to the right ($\gamma_{1.01}$) in Panel B. *T* values in parentheses. ***, ** and * denote significance at the 1, 5 and 10% levels, respectively (two-tailed)

are similarly developed countries (as measured by GDP per capita)²⁷ located in Western Europe, do not differ substantially in terms of the level of trust in public institutions (as measured by the corruption perception index)²⁸ and have similar tax rates in terms of CIT, PIT and VAT rates.²⁹ However, the specific design of the threshold-dependent policies differs strongly between the two countries. Specifically, whereas the German regime relies on multiple criteria for segmentation, the Spanish regime is based on a

²⁷ In 2020, Germany's GDP per capita was approximately 45,723 USD and Spain's GDP per capita was approximately 27,057 USD (The World Bank, 2021).

²⁸ According to the level of perceived public sector corruption (Transparency International, 2019), which can be used as a proxy for trust in public institutions, both German (global rank 9) and Spanish (global rank 32) institutions enjoy a high level of trust.

²⁹ The combined, i.e., including sub-central taxes, statutory CIT rate was 29.4% in Germany and 25% in Spain in 2010 (OECD, 2021a), the top statutory PIT rate was 43.5–47.5% (OECD, 2021b), and the standard VAT rate was 19–21% (OECD, 2020), respectively.

Table 7 Results for H1, H2, H3 (per AIC): *REV*

Panel A: Full sample						
FST	Left			Right		
	$\gamma_{0.95}$	$\gamma_{0.97}$	$\gamma_{0.99}$	$\gamma_{1.01}$	$\gamma_{1.03}$	$\gamma_{1.05}$
VSS	−0.0238 (−0.68)	−0.0118 (−0.34)	0.0217 (0.62)	0.0082 (0.24)	−0.0161 (−0.46)	−0.0447 (−1.28)
SM	0.0286 (1.37)	0.0099 (0.48)	−0.0192 (−0.92)	−0.0552** (−2.64)	−0.0215 (−1.03)	−0.0101 (−0.49)
ML	−0.0332 (−0.72)	0.0170 (0.37)	0.0401 (0.87)	−0.0234 (−0.51)	0.0071 (0.15)	0.0076 (0.16)
Panel B: per AIC						
AIC	VSS-FST		SM-FST		ML-FST	
	$\gamma_{0.99}$	$\gamma_{1.01}$	$\gamma_{0.99}$	$\gamma_{1.01}$	$\gamma_{0.99}$	$\gamma_{1.01}$
Trading	0.0179 (0.36)	0.0161 (0.33)	0.0490 (0.85)	−0.1450** (−2.52)	0.0813 (0.97)	−0.0891 (−1.07)
Manufacturing	0.1120** (2.21)	0.0008 (0.01)	−0.0049 (−0.16)	−0.0342 (−1.10)	0.0354 (0.39)	−0.0357 (−0.40)
Freelancing	−0.0700 (−1.02)	0.0901 (1.31)	−0.1020 (−1.15)	−0.0891 (−1.01)	0.158 (0.69)	0.378 (1.65)
Services	−0.0091 (−0.21)	−0.0008 (−0.02)	−0.0582 (−1.48)	−0.0159 (−0.40)	−0.0064 (−0.07)	−0.0262 (−0.29)

This table reports the regression coefficients from estimating Equation 1 for *REV*. Panel A presents findings for H1 and H2, i.e., results for the full sample of firms at the VSS-FST, SM-FST and the ML-FST for the segmentation cycle starting in 2010. Panel B presents the findings for H3, i.e., subsample results at the VSS-FST, SM-FST and ML-FST per AIC. We report the coefficients for all three bins to the left of the FST ($\gamma_{0.95}$, $\gamma_{0.97}$, $\gamma_{0.99}$) and all three bins to the right ($\gamma_{1.01}$, $\gamma_{1.03}$, $\gamma_{1.05}$) in Panel A but only the coefficient of the first bin to the left ($\gamma_{0.99}$) and the first bin to the right ($\gamma_{1.01}$) in Panel B. *T* values in parentheses. ***, ** and * denote significance at the 1, 5 and 10% levels, respectively (two-tailed)

single criterion. Furthermore, the German regime is generally more complex because it relies on four different size classes, regular adjustments of FSTs and industry-specific FSTs. By contrast, the Spanish regime only differentiates between two size classes, FSTs are fixed in nominal terms, and FSTs do not differ across industries. Hence, we argue that the specific design of threshold-dependent policies can inhibit size management by increasing firms' optimization costs. However, ultimately, we do not have a clear enough setting to provide direct evidence that the different outcomes for Spain and Germany are driven by the specific design of the threshold-dependent enforcement regime.

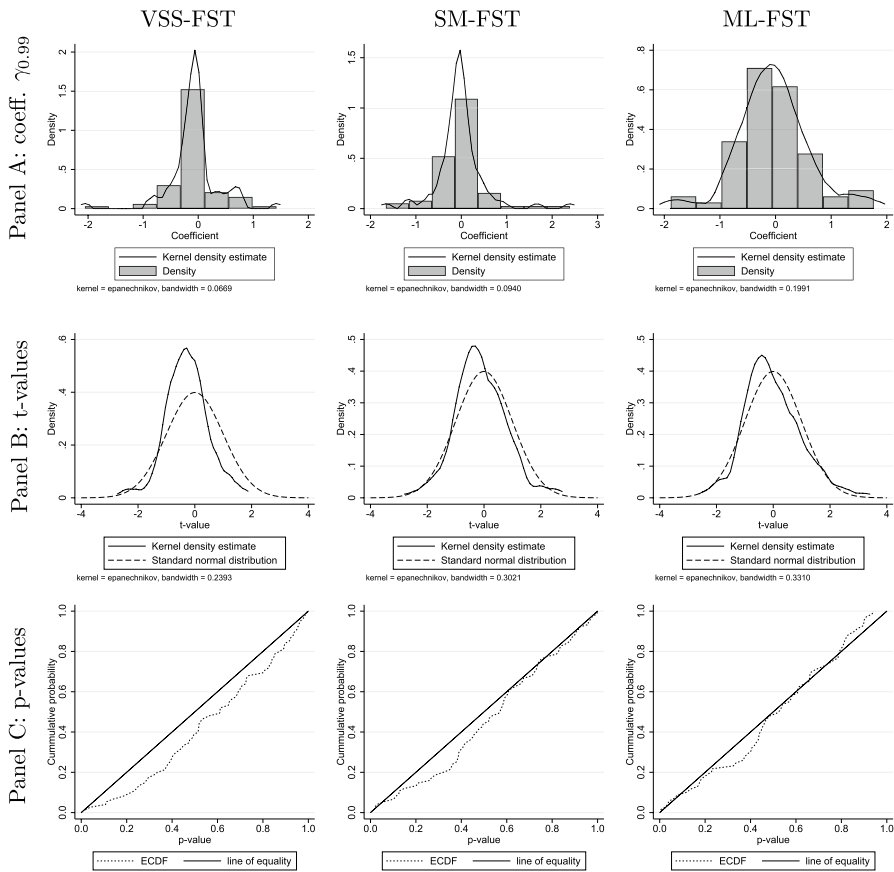


Fig. 4 Results for H3 (per industry): EBT. *Notes:* This figure graphs the distributions of regression coefficients ($\gamma_{0.99}$), t values and p values from estimating Eq. 1 at the VSS-FST, SM-FST and ML-FST for EBT for the segmentation cycle starting in 2010. Panel A: Histograms and kernel density estimates (solid line) for regression coefficients; Panel B: kernel density estimates (solid line) for t values and standard normal density distribution (dashed line); Panel C: ECDFs of p values (short-dashed line) and line of equality (solid line diagonal). The vertical axis presents the (empirical) density in Panel A and Panel B and the (empirical) cumulative probability in Panel C. The horizontal axis shows the coefficients, t values and p values, respectively

8 Robustness tests

8.1 Adjustment costs vs. information costs

In our setting, it is not possible to empirically disentangle the effects of the two components of optimization costs, i.e., adjustment costs and information costs. However, an absence of size management would be unlikely if adjustment costs were the only friction at work (Bosch et al., 2019; Søggaard, 2019). In particular, due to variable adjustment costs, it appears unlikely that adjustment costs exceed the decrease in expected firm audit costs for firms in close proximity to

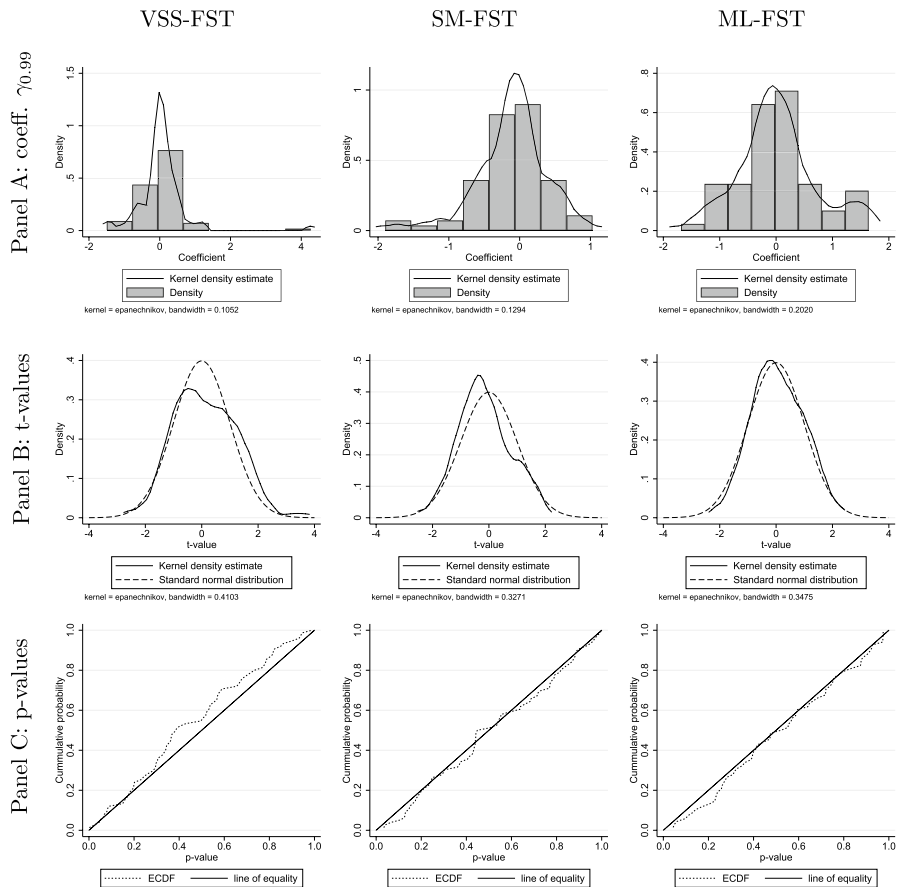


Fig. 5 Results for H3 (per industry): *REV*. *Notes:* This figure graphs the distributions of regression coefficients ($\gamma_{0,99}$), t values and p values from estimating Eq. 1 at the VSS-FST, SM-FST and ML-FST for *REV* for the segmentation cycle starting in 2010. Panel A: Histograms and kernel density estimates (solid line) for regression coefficients; Panel B: kernel density estimates (solid line) for t values and standard normal density distribution (dashed line); Panel C: ECDFs of p values (short-dashed line) and line of equality (solid line diagonal). The vertical axis presents the (empirical) density in Panel A and Panel B and the (empirical) cumulative probability in Panel C. The horizontal axis shows the coefficients, t values and p values, respectively

the FST. Therefore, we argue that information costs play an important role in our setting. To provide some evidence for this argument, we consider an additional setting in which bunching has been identified by prior studies. Specifically, we analyze the distribution of the financial accounting after-tax profits (as reported in CIT returns) around zero, as there is ample empirical evidence (Bollen and Pool, 2009; Burgstahler and Dichev, 1997; Lahr, 2014) that firms attempt to avoid reporting losses for various reasons. The histogram in Fig. 6 shows the distribution of firms' ratios of after-tax profits to *REV* around zero (solid vertical line) in a range between -0.2 and 0.2% . The bin width is set to 0.01% . The bunching

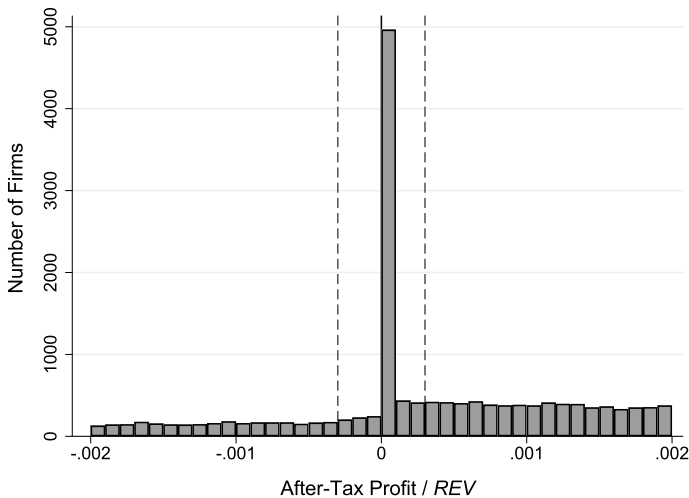


Fig. 6 Size management to avoid reporting losses. *Notes:* This figure graphs the distribution of firms' ratios of after-tax profits to *REV* around zero (solid line) in 2010. We focus on in the interval $[-0.2, 0.2\%]$. The bin width is 0.01%. The bunching interval is set to three bins to the right and three bins to the left (dashed vertical lines)

interval is set to three bins to the right and three bins to the left (dashed vertical lines).

There is a discernible discontinuity in the distribution of firms at zero in the otherwise smooth (uniform) distribution, i.e., there is bunching above zero.³⁰ Furthermore, we estimate Equation 1 at zero for firms with after-tax profitability between -0.2 and 0.2% and the bin width set to 0.01% . All three regression coefficients to the right are significantly positive (not tabulated), which implies that there is an excess mass between zero and 0.03% . The first coefficient to the right is significantly larger than the second coefficient and the third coefficient, which implies that firms prefer to manage their size by the smallest amount necessary to exceed the implicit threshold, suggesting variable adjustment costs. The coefficients to the left are negative but nonsignificant, suggesting that size-managing firms originate from a large area below the threshold, i.e., the missing mass is rather dispersed. Considered jointly, the results provide some indication that firms in our data practice size management and hence that adjustment costs are unlikely the only friction at work. In addition, the results provide some evidence for the sensitivity of our test to detect bunching.

8.2 Time effects

Specific time effects might have prevented size management in 2010. One reason for such an effect could be, among others, the financial crisis around that time. To

³⁰ As firms with missing *REV* are excluded, the results are not driven by inactive firms that naturally report zero profits.

ensure that our results are not only prevalent in 2010, we repeat our baseline analyses from Tables 6 and 7 as well as Panel C of Figs. 4 and 5 using data for 2004 and 2007 (not tabulated or graphed).³¹ We again do not find any evidence of size management around FSTs.

8.3 Firms not exceeding the respective other firm size threshold

Due to variable adjustment costs firms exceeding the FST for revenue by far and thus facing adjustment costs that exceed the benefits of size management have no incentive to manage size at the respective FST for profit and vice versa. To reduce noise in our analyses that might stem from keeping such firms in the sample, we repeat the baseline analyses from Tables 6 and 7 as well as Panel C of Figs. 4 and 5 while restricting our sample to firms that do not exceed the respective FSTs for revenue (profit) when examining the FSTs for profit (revenue) (not tabulated or graphed). However, the results remain virtually unchanged.

8.4 Loss firms

Chen and Lai (2012), Edwards et al. (2016) and Law and Mills (2015) show that due to a higher cost of external financing financially constrained firms engage in more aggressive tax avoidance than unconstrained firms to increase internally generated funds. Correspondingly, loss firms might have larger incentives to engage in size management at FSTs for revenue. Hence, we repeat our baseline analyses from Table 7 while restricting our sample to firms with negative *EBT* (not tabulated or graphed). However, we again do not find any evidence of size management around FSTs.

8.5 Geographic heterogeneity

Audit intensity may vary between German states due to different resources being available for audits (see Sect. 3.1). Hence, it is conceivable that size management occurs only in states that allocate substantial resources to audits and that the respective effects in our full sample analysis are covered by the noise of states without effects. We therefore repeat the analyses from Panel B of Tables 6 and 7 per state instead of per AIC (not tabulated). However, we do not find any evidence of size management around FSTs, suggesting an absence of size management for all 16 states.

Along the same lines, as audits are conducted by local tax offices, audit intensity can also be conditional on the specific tax office responsible for an audit. Each tax office is usually responsible for one of the 400 German districts. Hence, it is feasible that size management is heterogeneous across individual districts. Therefore, we

³¹ We obtain the exact same data for 2004 and 2007 as for 2010. As firm identifiers and firm names are not included in the data, it is, however, not possible to merge observations over time.

replicate the baseline analyses from Panel C of Figs. 4 and 5 (not graphed) per district instead of per industry. We again do not find any evidence of size management around FSTs.

8.6 Relevant firm size thresholds

Due to marginal adjustments of FSTs before each segmentation cycle, firms do not know the exact FSTs that will be applied in the next segmentation cycle when they have to engage in size management (see Sect. 3.2). However, firms are aware of FSTs applied for the current segmentation cycle when they have to engage in size management, and FSTs have historically never decreased. Consequently, we assume in our baseline analyses that firms using a conservative approach manage their size to the FSTs applied for the current segmentation cycle. However, some firms could also be less risk averse and attempt to predict the FSTs that will be applied in the next segmentation cycle, and hence, these firms would bunch in an area above the FSTs applied in the last segmentation cycle. If this is the case, the baseline analyses would not be well suited to detect size management. Therefore, we repeat the baseline analyses from Panel A of Tables 6 and 7 at different placebo FSTs (not tabulated). To obtain the placebo FSTs, we start with the FSTs applied for the segmentation cycle starting in 2010 and gradually increase FSTs in steps of 100 euros until the placebo FSTs correspond to the FSTs applied for the segmentation cycle starting in 2013. However, we still do not find any evidence of size management around those placebo FSTs.³²

9 Conclusion

This paper contributes to the recent literature on the effects of threshold-dependent tax enforcement. We analyze the response of German firms to discontinuities in audit intensity at publicly known FSTs. Given that tax audits usually result in substantial tax claims, interest payments and penalty fees and can cause substantial compliance costs, it would be expected that size management occurs around the FSTs. Using a large administrative dataset of tax returns, we test this prediction and exploit discontinuities in the firm size distribution that would be expected from size management. Building on established tests for bunching in the context of notches (Chetty et al., 2011; Kleven and Waseem, 2013; Saez, 2010), our empirical results indicate that there is no tax-induced size management in the overall population of

³² Alternatively, the test developed by Ullmann and Watrin (2017) might provide a suitable empirical strategy when the exact FST that firms chose for their size management is unknown. The test does not require information on exact target values and instead relies on the concept of the distribution of digits rather than the distribution of the size variable itself. However, as the test does not rely on a theoretically derived distribution but relative comparisons of the distributions of digits, the test requires data on at least two groups of firms, where at least one group has to have unmanaged size variables. Such an unmanaged group is not available in our setting because even FSTs for different AICs are relatively close to each other.

German firms. The results hold when excessive testing in a large variety of subsamples is conducted, when alternative bunching tests are applied and when different alternative periods of analysis and alternative FSTs are used.

We posit that the absence of size management results from optimization costs in the form of adjustment costs and information costs. Against the background of prior research, we argue that a pattern seems to be emerging that the specific design of threshold-dependent policies can inhibit size management. Specifically, we argue that using multiple criteria for segmentation, multiple size classes, regular adjustments of FSTs after firm decisions are made and industry-specific FSTs increase optimization costs and, hence, can inhibit size management. Therefore, our findings provide relevant implications for policy makers, as they suggest that the specific design of threshold-dependent policies might allow governments to increase the efficiency of tax audits without distorting the firm size distribution and, hence, avoid the negative effects of size management on welfare. However, more research is needed to granularly disentangle the effects that individual characteristics of threshold-dependent enforcement regimes have on optimization costs.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10797-022-09732-2>.

Acknowledgement We thank Nadine Riedel (the editor), two anonymous reviewers, Antonio de Vito (discussant), Rainer Niemann (discussant), Lisa Hillmann (discussant), as well as participants at the 5th Berlin-Vallendar Conference, the 6th Annual MaTax Conference, the 11th Norwegian-German Seminar, the 7th Augolstadt Seminar, the 82nd VHB Annual Conference for helpful comments. We are indebted to Melanie Heiliger and Anette Erbe (both RDC of the Federal Statistical Office and the Statistical Offices of the Federal States) for their invaluable support in facilitating remote analysis of the confidential tax return data.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability RDC of the Federal Statistical Office and the Statistical Offices of the Federal States, Integrated Tax Return Data, 2004 (DOI: 10.21242/73511.2004.00.04.2.1.0), 2007 (DOI: 10.21242/73511.2007.00.04.2.1.0) and 2010 (DOI: 10.21242/73511.2010.00.04.2.1.0).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agostini, C. A., Engel, E., Repetto, A., & Vergara, D. (2018). Using small businesses for individual tax planning: Evidence from special tax regimes in Chile. *International Tax and Public Finance*, 25(6), 1449–1489.

- Alm, J. (2019). What Motivates Tax Compliance? *Journal of Economic Surveys*, 33(2), 353–388.
- Alm, J., Jackson, B. R., & McKee, M. (2009). Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics*, 93(3–4), 392–402.
- 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.
- Asatryan, Z., & Peichl, A. (2017). Responses of firms to tax, administrative and accounting rules: Evidence from Armenia. *Working Paper*.
- Asatryan, Z., Peichl, A., Schwab, T., Voget, J. (2018). Inverse December fever. *Working Paper*.
- Bachas, P., Fattal Jaef, R. N., & Jensen, A. (2019). Size-dependent tax enforcement and compliance: Global evidence and aggregate implications. *Journal of Development Economics*, 140, 203–222.
- Bachas, P., & Soto, M. (2021). Corporate taxation under weak enforcement. *American Economic Journal: Economic Policy*, 13(4), 36–71.
- Bavarian General Accounting Office (2013). Annual Report 2013 [Jahresbericht 2013: TNr. 19: Betriebssprüfung stärken].
- Bernard, D., Burgstahler, D., & Kaya, D. (2018). Size management by European private firms to minimize proprietary costs of disclosure. *Journal of Accounting and Economics*, 66(1), 94–122.
- Best, M. C., Brockmeyer, A., Kleven, H. J., Spinnewijn, J., & Waseem, M. (2015). Production versus revenue efficiency with limited tax capacity: Theory and evidence from Pakistan. *Journal of Political Economy*, 123(6), 1311–1355.
- Bollen, N. P., & Pool, V. K. (2009). Do hedge fund managers misreport returns? Evidence from the pooled distribution. *The Journal of Finance*, 64(5), 2257–2288.
- Bosch, N., Jongen, E., Leenders, W., & Möhlmann, J. (2019). Non-bunching at kinks and notches in cash transfers in the Netherlands. *International Tax and Public Finance*, 26(6), 1329–1352.
- Brockmeyer, A. (2014). The investment effect of taxation: Evidence from a corporate tax kink. *Fiscal Studies*, 35(4), 477–509.
- Brown, C. V. (1969). Misconceptions about income tax and incentives. *Scottish Journal of Political Economy*, 16(2), 1–21.
- Burgstahler, D., & Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics*, 24(1), 99–126.
- Chen, C., & Lai, S. (2012). Financial constraint and tax aggressiveness. *Working Paper*.
- Chen, S., Chen, X., Cheng, Q., & Shevlin, T. J. (2010). Are family firms more tax aggressive than non-family firms? *Journal of Financial Economics*, 95(1), 41–61.
- Chetty, R., Friedman, J. N., Olsen, T., & Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *The Quarterly Journal of Economics*, 126(2), 749–804.
- Devereux, M. P., Liu, L., & Loretz, S. (2014). The elasticity of corporate taxable income: New evidence from UK tax records. *American Economic Journal: Economic Policy*, 6(2), 19–53.
- Edwards, A., Schwab, C., & Shevlin, T. (2016). Financial constraints and cash tax savings. *The Accounting Review*, 91(3), 859–881.
- Fujii, E. T., & Hawley, C. B. (1988). On the accuracy of tax perceptions. *The Review of Economics and Statistics*, 70(2), 344–347.
- Gallimore, J., & Jacob, M. (2020). Corporate tax enforcement externalities and the banking sector. *Journal of Accounting Research*, 58(5), 1117–1159.
- Gallimore, J., & Labro, E. (2015). The importance of the internal information environment for tax avoidance. *Journal of Accounting and Economics*, 60(1), 149–167.
- Garicano, L., Lelarge, C., & van Reenen, J. (2016). Firm size distortions and the productivity distribution: Evidence from France. *American Economic Review*, 106(11), 3439–3479.
- German Bundestag (2021). Document No. 19/29616 [Drucksache 19/29616: Fallauswahl im Rahmen der Außenprüfung durch die Finanzbehörden].
- German Federal Court of Finance (1988). Tax audit of small and medium firms [Außenprüfung bei Mittel- und Kleinbetrieben, Az. III R 280/84].
- German Federal Ministry of Finance (2003). Tax audit thresholds 2004 [Schreiben betr. Einordnung in Größenklassen gem. § 3 BpO 2000; Merkmale zum 1. Januar 2004].
- German Federal Ministry of Finance (2005). Tax audit results 2004 [Ergebnisse der steuerlichen Betriebsprüfung 2004].
- German Federal Ministry of Finance (2006). Tax audit thresholds 2007 [Schreiben betr. Einordnung in Größenklassen gem. § 3 BpO 2000; Festlegung neuer Merkmale zum 1. Januar 2007].

- German Federal Ministry of Finance (2008). Tax audit results 2007 [Ergebnisse der steuerlichen Betriebsprüfung 2007].
- German Federal Ministry of Finance (2009). Tax audit thresholds 2010 [Schreiben betr. Einordnung in Größenklassen gem. §3 BpO 2000; Festlegung neuer Abgrenzungsmerkmale zum 1. Januar 2010].
- German Federal Ministry of Finance (2011). Tax audit results 2010 [Ergebnisse der steuerlichen Betriebsprüfung 2010].
- Graham, J. R., Hanlon, M., Shevlin, T., & Shroff, N. (2017). Tax rates and corporate decision-making. *The Review of Financial Studies*, 30(9), 3128–3175.
- Guedhami, O., & Pittman, J. (2008). The importance of IRS monitoring to debt pricing in private firms. *Journal of Financial Economics*, 90(1), 38–58.
- Gupta, S., & Newberry, K. (1997). Determinants of the variability in corporate effective tax rates: Evidence from longitudinal data. *Journal of Accounting and Public Policy*, 16(1), 1–34.
- Harju, J., Matikka, T., & Rauhanen, T. (2016). The effects of size-based regulation on small firms: Evidence from VAT threshold. *Working Paper*.
- Harle, G., & Olles, U. (2017). *Modern Tax Audits [Die moderne Betriebsprüfung]*. NWB, Herne, 3rd edition.
- Henselmann, K., & Haller, S. (2017). Potential risk factors for the increase of the tax audit probability [Potentielle Risikofaktoren für die Erhöhung der Betriebsprüfungswahrscheinlichkeit-Eine analytische und empirische Untersuchung auf Basis der E-Bilanz-Taxonomie 6.0]. *Working Paper*.
- Hoopes, J. L., Mescall, D., & Pittman, J. A. (2012). Do IRS audits deter corporate tax avoidance? *The Accounting Review*, 87(5), 1603–1639.
- Hoopes, J. L., Reck, D. H., & Slemrod, J. (2015). Taxpayer search for information: Implications for rational attention. *American Economic Journal: Economic Policy*, 7(3), 177–208.
- Hoopes, J. L., Robinson, L., & Slemrod, J. (2018). Public tax-return disclosure. *Journal of Accounting and Economics*, 66(1), 142–162.
- Hosono, K., Hotei, M., Miyakawa, D. (2018). Tax avoidance by capital reduction: Evidence from corporate tax reform in Japan. *Working Paper*.
- Kaligin, T. (2014). *Tax Audits and Tax Investigation [Betriebsprüfung und Steuerfahndung]*. Stuttgart: Boorberg.
- Kanbur, R., & Keen, M. (2014). Thresholds, informality, and partitions of compliance. *International Tax and Public Finance*, 21(4), 536–559.
- Klein, F., & Rüskens, R., editors (2020). *German Tax Regulation [AO §194 Rn. 20-22]*. 15th edition.
- Kleven, H. J., & Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *The Quarterly Journal of Economics*, 128(2), 669–723.
- Kosonen, T., & Matikka, T. (2019). Discrete earnings responses to tax incentives: Empirical evidence and implications. *Working Paper*.
- Kreditanstalt für Wiederaufbau (2012). SME Panel [KfW-Mittelstandspanel].
- Lahr, H. (2014). An improved test for earnings management using kernel density estimation. *European Accounting Review*, 23(4), 559–591.
- Law, K. K. F., & Mills, L. F. (2015). Taxes and financial constraints: Evidence from linguistic cues. *Journal of Accounting Research*, 53(4), 777–819.
- Liu, L., Lockwood, B., Almunia, M., Tam, E.H. (2019). VAT notches, voluntary registration, and bunching: Theory and UK evidence. *Working Paper*.
- Meyer, H. (1988). Tax audit consulting fees [Die Vergütung für eine steuerliche Betriebsprüfung]. *KP Kanzleiführung professionell*, (07/1998):4.
- OECD (2015). Tax administration 2015: Comparative information on OECD and other advanced and emerging economies. Paris.
- OECD (2017). Tax administration 2017: Comparative information on OECD and other advanced and emerging economies. Paris.
- OECD (2020). Consumption tax trends 2020: VAT/GST and excise rates, trends and policy issues.
- OECD (2021a). OECD.Stat: Tax database: Statutory corporate income tax rate.
- OECD (2021b). OECD.Stat: Tax database: Top statutory personal income tax rates.
- Onji, K. (2009). The response of firms to eligibility thresholds: Evidence from the Japanese value-added tax. *Journal of Public Economics*, 93(5–6), 766–775.
- Panek, M. (2018). Case selection and determination of audit focal points for tax audits [Fallauswahl und Festlegung von Prüfungsschwerpunkten für die Betriebsprüfung]. *Der Betrieb*, 71(Supplement 2/2018), 31–35.

- PricewaterhouseCoopers (2019). Tax audits 2018 [Betriebsprüfung 2018 – Studie zur Praxis der Betriebsprüfung in Deutschland].
- Regional Tax Authority of Rhineland-Palatinate (2016). Annual report 2015 [Jahresbericht 2015].
- Roychowdhury, S. (2006). Earnings management through real activities manipulation. *Journal of Accounting and Economics*, 42(3), 335–370.
- Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3), 180–212.
- Slemrod, J. (2016). Tax compliance and enforcement: New research and its policy implications. *Working Paper*.
- Søgaard, J. E. (2019). Labor supply and optimization frictions: Evidence from the Danish student labor market. *Journal of Public Economics*, 173, 125–138.
- Strangmeier, R. (2000). *The tax audit and the uncertainty of economic success [Die steuerliche Betriebsprüfung und die Unbestimmtheit des ökonomischen Erfolges: Eine wirtschaftssoziologische Studie mit einer Analyse der Groß- und Konzernbetriebsprüfung]*. Bielefeld: Erich Schmidt.
- Tennant, S. N., & Tracey, M. R. (2019). Corporate profitability and effective tax rate: The enforcement effect of large taxpayer units. *Accounting and Business Research*, 49(3), 342–361.
- The World Bank. (2017). Doing business 2017: Equal opportunity for all.
- The World Bank. (2021). National accounts data: GDP per capita.
- Transparency International. (2019). Corruption perceptions index 2018.
- Ullmann, R., & Watrin, C. (2017). Detecting target-driven earnings management based on the distribution of digits. *Journal of Business Finance and Accounting*, 44(1–2), 63–93.
- Vehorn, C. L. (2011). Fiscal adjustment in developing countries through tax administration reform. *The Journal of Developing Areas*, 45(1), 323–338.
- Wenzig, H. (2014). *Tax Audits [Außenprüfung/Betriebsprüfung]*. Grüne Reihe / Steuerrecht für Studium und Praxis. Erich Fleischer Verlag, Achim, 10th edition.
- Zimmerman, J. L. (1983). Taxes and firm size. *Journal of Accounting and Economics*, 5, 119–149.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.