Three Essays on The Effects of Taxes on Individuals, Firms and Governments

Kumulative Dissertation

der Wirtschaftswissenschaftlichen Fakultät

der Universität Augsburg

zur Erlangung des Grades eines

Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

vorgelegt von

Drahomir Klimsa

(M. Sc.)

Mai 2022

Erstgutachter:	Prof. Dr. Robert Ullmann
Zweitgutachter:	Prof. Dr. Marco Wilkens
Vorsitzender der mündlichen Prüfung:	Prof. Dr. Erik Lehmann
Tag der mündlichen Prüfung:	28.7.2022

List of Essays

This doctoral thesis contains the following three essays:

- **Essay 1: Threshold-Dependent Tax Enforcement and the Size Distribution of Firms: Evidence from Germany, *International Tax and Public Finance*, Published 22 April 2022, jointly with Robert Ullmann.**

  Accepted for presentation at the 5th Berlin-Vallendar Conference (Berlin), the 6th Annual MaTax Conference (Mannheim), the 11th Norwegian-German Seminar (Munich), and the 82nd VHB Annual Conference (online).

- **Essay 2: Norderfriedrichskoog! Tax Havens, Tax Competition and the Introduction of a Minimum Tax Rate, *Working Paper*, jointly with William C. Boning, Joel Slemrod, and Robert Ullmann.**

  Accepted for presentation at the 2018 Oxford University Centre of Business Taxation Academic Symposium (Oxford), the 5th Annual MaTax Conference (Mannheim), and the 23rd Annual Conference on Finance and Accounting (Prague).

- **Essay 3: The Effects of Scandals on Organizational Affiliation and Competition: Evidence from Church Scandals in Germany, *Working Paper*, jointly with Erik E. Lehmann, Robert Ullmann, and Laurenz Weiße.**

  Accepted for presentation at the 44th Annual Congress of the European Accounting Association (Bergen), the 38th European Group for Organizational Studies Colloquium (Wien), and the 82nd Annual Meeting of the Academy of Management (Seattle).

Abstracts

Essay 1: Threshold-Dependent Tax Enforcement and the Size Distribution of Firms: Evidence from Germany

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

Essay 2: Norderfriedrichskoog! Tax Havens, Tax Competition and the Introduction of a Minimum Tax Rate

German municipalities levy local business taxes (Gewerbesteuer) by choosing a tax rate to apply to local reported business income, where the tax base is defined uniformly at the national level. Prior to the federal government's imposition of a minimum tax rate in 2004, some municipalities, such as the tiny North Sea town of Norderfriedrichskoog, chose to act as tax havens by setting a zero tax rate. We combine administrative microdata from firm tax returns with municipality-level information to study the choice of becoming a tax haven; the extent to which havens attracted income from other municipalities before and after the introduction of the minimum tax rate; and how the introduction of the minimum tax rate affected the tax competition equilibrium among non-haven municipalities. Our results suggest that income was shifted to haven municipalities both before and after the introduction of the minimum tax rate. Our findings also indicate that the mandated increase in havens' tax rates did not lead to rate increases (or decreases) among municipalities in general or tax-haven municipalities' geographical neighbors. In contrast to the literature on global business tax competition, our preferred specifications, which leverage the minimum tax rate imposition for identification, find no evidence of competition in business tax rates. We find that tax havens largely do not affect the business tax rates set by non-havens, suggesting that a global minimum tax rate binding only for international tax havens will have little effect on tax competition between non-haven countries.

Keywords: income shifting, minimum tax, tax competition, tax havens

JEL Classification: H26, H32, H71

Essay 3: The Effects of Scandals on Organizational Affiliation and Competition: Evidence from Church Scandals in Germany

This paper investigates the effects of scandals on organizations and their stakeholders. We introduce a novel framework that links the conceptual origin of a scandal, i.e., individual-caused vs. institution-caused, with its impact on affiliation with the scandal-stricken organization and with the scandal-stricken organization's competitors. In our analysis, we exploit the changes in diocese-level and regional church-level measures of affiliation with the Catholic Church and Protestant Church in Germany that followed numerous scandals involving the two major German religious organizations between 2002 and 2016. We find that both individual-caused and institution-caused scandals are associated with a decline in affiliation with the scandal-stricken organization. However, individual-caused scandals have a significantly larger effect on affiliation with the scandal-stricken organization than institution-caused scandals. We also find evidence of positive interorganizational spillover effects on unassociated competitors of the scandal-stricken organization, but only for institution-caused scandals. Our results contribute to the emerging field of studies on the effects of scandals on organizations and their stakeholders. Moreover, due to the economic character of religious organizations, i.e., because they compete in a religious market to provide pastoral care, our results can be generalized beyond our empirical setting to secular organizations and their stakeholders.

Keywords: church tax, organizations, religion, reputation, scandals

JEL Classification: J11, L14, L30, M14, Z12

Introduction

1. Motivation

Tax research has a long history and has been conducted across various disciplines, including accounting, finance, economics, and law. In recent years, a substantial part of tax research, particularly empirical tax research, has focused on how tax policies affect real economic decisions, i.e., the observable responses of the economic agents affected by such policies (Hanlon and Heitzman, 2010). According to Shackelford and Shevlin (2001), the main topics addressed by this microeconomics-based literature can be summarized into three fundamental questions:

1. "*Do taxes matter?*"
2. "*If not, why not?*"
3. "*If so, how much?*"

Most empirical tax research builds on the Scholes and Wolfson framework (Hanlon and Heitzman, 2010). Once developed in the context of corporate tax planning,[1] the framework, which is based on the three central themes "*all parties*," "*all taxes*," and "*all costs*," has become a key tool in the analysis of the effects of tax policies in empirical tax research (Shackelford and Shevlin, 2001; Shevlin, 2020). The implicit assumption of the framework is that if all concerned parties, all taxes (explicit and implicit) and all nontax costs can be identified and controlled for, an economic agent's response to a specific tax policy is rational and predictable (Shackelford and Shevlin, 2001).

Despite a large body of research,[2] there are still many questions regarding the effects of tax policies on economic agents that remain unanswered. The three essays included in this doctoral thesis aim to close some of the gaps in the literature by analyzing the effects of tax policies in three understudied, yet important areas of taxation.

1 The framework was first illustrated in Scholes et al. (1990).
2 For a comprehensive discussion of the prior literature, see the seminal papers by Hanlon and Heitzman (2010), Maydew (2001), and Shackelford and Shevlin (2001).

Figure 1: Areas of Taxation Addressed by the Essays Included in the Doctoral Thesis

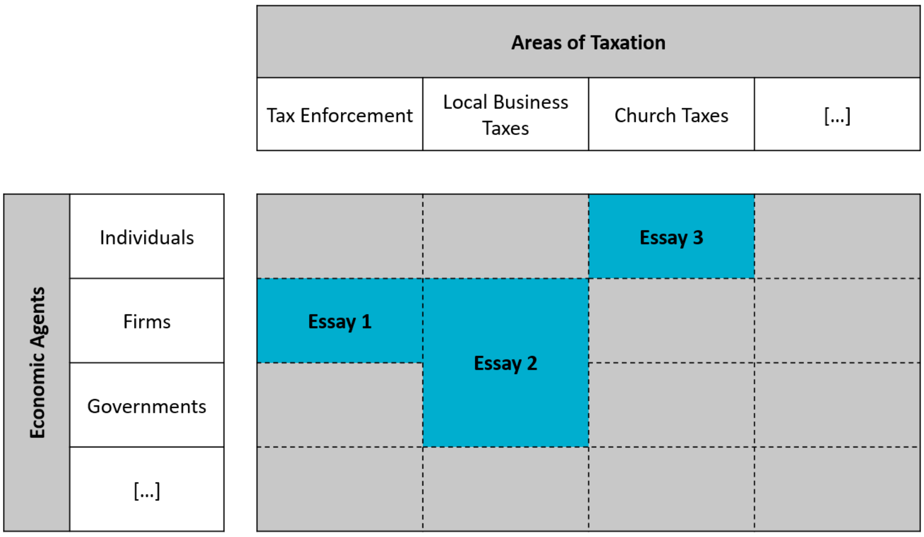


Figure 1 provides an overview of the areas of taxation addressed by the three essays included in this doctoral thesis. On the vertical axis, the essays are classified by the respective economic agents of interest. Specifically, I differentiate whether the response of a) individuals, b) firms or c) governments is analyzed. On the horizontal axis, the essays are classified by the respective area of taxation where the effects of specific tax policies are studied, i.e., a) tax enforcement, b) local business taxes, and c) church taxes.

Whereas the first essay *"Threshold-Dependent Tax Enforcement and the Size Distribution of Firms: Evidence from Germany"* (Essay 1) studies how firms respond to a widely applied type of tax enforcement policy, specifically, threshold-dependent tax enforcement, the second essay *"Norderfriedrichskoog! Tax Havens, Tax Competition and the Introduction of a Minimum Tax Rate"* (Essay 2) studies intrastate tax competition in local business taxes, i.e., a specific type of income tax set by local governments, and profit shifting of firms to domestic tax havens. The third essay *"The Effects of Scandals on Organizational Affiliation and Competition: Evidence from Church Scandals in Germany"* (Essay 3) examines the response of individuals to scandals related to organizations in a specific setting where religious organizations levy taxes, namely, so-called church taxes, on their members.

The remainder of this introduction provides a brief review of prior literature in the addressed areas of taxation, summarizes the three essays at hand and outlines their contribution to the literature.

2. Summary and Contribution to the Literature

2.1. Essay 1: Threshold-Dependent Tax Enforcement and the Size Distribution of Firms: Evidence from Germany

Despite the worldwide adoption of threshold-dependent tax enforcement policies (approximately 85% of the world's 60 largest economies have adopted such policies (OECD, 2015)), research on this subject is scarce. Apart from Essay 1, there are only two studies (Almunia and Lopez-Rodriguez, 2018; Tennant and Tracey, 2019) that analyze how firms respond to threshold-dependent tax enforcement. However, prior research has revealed size management of firms and individuals at publicly known thresholds in different areas of taxation, i.e., public tax return disclosure (Hoopes et al., 2018), kinks and notches in corporate income tax rates (Bachas and Soto, 2021; Brockmeyer, 2014; Devereux et al., 2014), tax benefits and special tax regimes for SMEs (Agostini et al., 2018; Hosono et al., 2018), minimum tax schemes (Best et al., 2015) and value added tax exemptions (Onji, 2009; Liu et al., 2019; Harju et al., 2016). Moreover, size management has also been documented with respect to an array of nontax regulations, e.g., mandatory IFRS reporting (Asatryan and Peichl, 2017), financial audit and disclosure requirements (Bernard et al., 2018) and labor regulations (Garicano et al., 2016).

Essay 1 contributes to this recent literature by investigating firms' responses to threshold-dependent intensity of tax enforcement. Specifically, the study uses administrative tax return data on the entire population of German firms and exploits industry variation in firm size thresholds applied by the German tax administration. In the analyzed setting, each threshold marks a considerable spike in audit intensity and hence should create strong incentives for firms to bunch below the threshold. However, the large sample analysis finds no such effect in the observed setting. This empirical observation is attributed to optimization costs, particularly to the costs associated with the operational implementation of size management, i.e., adjustment costs, and the costs of gathering the relevant information about the threshold-dependent policy, i.e., information costs.

The study adds to the emerging field of studies on potential distortions created by threshold-dependent tax regulations. The findings are also relevant 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.

2.2. Essay 2: Norderfriedrichskoog! Tax Havens, Tax Competition and the Introduction of a Minimum Tax Rate

Although extensive literature is available on global income tax competition and particularly profit shifting to international tax havens (e.g., Dharmapala and Hines (2009), Dharmapala and Riedel (2013), Dharmapala (2014), Grubert and Slemrod (1998), Heckemeyer and Overesch (2017), Hines and Rice (1994), Slemrod (2008), and Tørsløv et al. (2018)), a much smaller body of literature examines tax competition at the level of local governments and profit shifting to domestic tax havens. Specifically, Fossen and Steiner (2018), Ilchmann et al. (2015), Langenmayr and Simmler (2021), and von Schwerin and Buettner (2016), study tax competition in the context of the German local business tax, relying mostly on aggregate tax budget data.

Essay 2 adds to this literature by studying intrastate tax competition between local governments and profit shifting of firms to domestic tax havens. German municipalities levy local business taxes (Gewerbesteuer) by choosing a tax rate to apply to local reported business income, where the tax base is defined uniformly at the national level. Prior to the federal government's imposition of a minimum tax rate in 2004, some municipalities, e.g., the North Sea town of Norderfriedrichskoog, chose to act as tax havens by setting a zero tax rate. The study combines administrative firm-level data from tax returns with municipality-level data to study municipalities' choice to become a tax haven, the extent to which tax havens attracted income from other municipalities before and after the introduction of the minimum tax rate, and the effect of the introduction of the minimum tax rate on the tax competition equilibrium among non-haven municipalities. The results suggest that income was shifted to haven municipalities both before and after the introduction of the minimum tax rate. Furthermore, the findings indicate that the mandated increase in havens' tax rates did not lead to rate increases (or decreases) among municipalities in general or tax-haven municipalities' geographical neighbors. In contrast to the literature on global business tax competition, the preferred specifications, which leverage the minimum tax rate imposition for identification, find no evidence of competition in business tax rates.

Against the background of the recent introduction of a global minimum corporate tax rate, the results are also relevant for policy-makers as they indicate that tax havens largely do not affect the business tax rates set by non-havens, suggesting that a global minimum tax rate binding only for international tax havens will have little effect on tax competition between non-haven countries.

2.3. *Essay 3: The Effects of Scandals on Organizational Affiliation and Competition: Evidence from Church Scandals in Germany*

Research in the area of church taxes is relatively scarce. However, a strand of literature on religious disaffiliation has identified church taxes as a significant motive in individuals' decisions to disaffiliate from a religious organization (Berghammer et al. (2017), Kühn (2015), and Riegel et al. (2019)). Another strand of literature in organizational and management studies investigates the effects of scandals on organizations and their stakeholders (Bottan and Perez-Truglia (2015), Frick et al. (2021), Frick and Simmons (2017), Hungerman (2013), and Piazza and Jourdan (2018)).

Essay 3 brings together these different strands of literature by studying the effects of scandals in a specific setting where religious organizations levy church taxes on their members which impose substantial costs on membership and hence represent a mediator for the potential effects of scandals on organizational affiliation and, due to spillover effects, on organizational competition. Specifically, the study exploits changes in diocese-level and regional church-level data on various measures of affiliation following numerous scandals that occurred in German Catholic dioceses and Protestant regional churches between 2002 and 2016. The results indicate that scandals are associated with a decline in affiliation with the scandal-stricken organization and positive spillover effects on affiliation with unassociated competitors of the scandal-stricken organization.

The study contributes to the research on religious disaffiliation and to the emerging field of studies on the effects of scandals on organizations and their stakeholders. Due to the economic character of religious organizations, i.e., because religious organizations compete on the religious market for pastoral care, the results can be generalized beyond the specific empirical setting, particularly to stakeholders of secular organizations.

References

Agostini, C. A., Engel, E., Repetto, A., and 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.

Almunia, M. and 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. and Peichl, A. (2017). Responses of Firms to Tax, Administrative and Accounting Rules: Evidence from Armenia. *Working Paper*.

Bachas, P. and Soto, M. (2021). Corporate Taxation under Weak Enforcement. *American Economic Journal: Economic Policy*, 13(4):36–71.

Berghammer, C., Zartler, U., and Krivanek, D. (2017). Looking Beyond the Church Tax: Families and the Disaffiliation of Austrian Roman Catholics. *Journal for the Scientific Study of Religion*, 56(3):514–535.

Bernard, D., Burgstahler, D., and 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., and Waseem, M. (2015). Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan. *Journal of Political Economy*, 123(6):1311–1355.

Bottan, N. L. and Perez-Truglia, R. (2015). Losing my religion: The effects of religious scandals on religious participation and charitable giving. *Journal of Public Economics*, 129:106–119.

Brockmeyer, A. (2014). The Investment Effect of Taxation: Evidence from a Corporate Tax Kink. *Fiscal Studies*, 35(4):477–509.

Devereux, M. P., Liu, L., and Loretz, S. (2014). The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records. *American Economic Journal: Economic Policy*, 6(2):19–53.

Dharmapala, D. (2014). What Do We Know about Base Erosion and Profit Shifting? A Review of the Empirical Literature. *Fiscal Studies*, 35(4):421–448.

Dharmapala, D. and Hines, J. R. (2009). Which countries become tax havens? *Journal of Public Economics*, 93(9):1058–1068.

Dharmapala, D. and Riedel, N. (2013). Earnings shocks and tax-motivated income-shifting: Evidence from European multinationals. *Journal of Public Economics*, 97:95–107.

Fossen, F. M. and Steiner, V. (2018). The Tax-rate Elasticity of Local Business Profits. *German Economic Review*, 19(2):162–189.

Frick, B., Moser, K., and Simmons, R. (2021). Spillover effects of scandals on exits from the catholic and protestant churches in germany. *Journal for the Scientific Study of Religion*, 60(3):482–497.

Frick, B. and Simmons, R. (2017). The impact of exogenous shocks on exits from the Catholic and Protestant churches in Germany, 1953–2015. *Applied Economics Letters*, 24(20):1476–1480.

Garicano, L., Lelarge, C., and van Reenen, J. (2016). Firm Size Distortions and the Productivity Distribution: Evidence from France. *American Economic Review*, 106(11):3439–3479.

Grubert, H. and Slemrod, J. (1998). The Effect of Taxes on Investment and Income Shifting to Puerto Rico. *The Review of Economics and Statistics*, 80(3):365–373.

Hanlon, M. and Heitzman, S. (2010). A review of tax research. *Journal of Accounting and Economics*, 50(2-3):127–178.

Harju, J., Matikka, T., and Rauhanen, T. (2016). The Effects of Size-Based Regulation on Small Firms: Evidence from VAT Threshold. *Working Paper*.

Heckemeyer, J. H. and Overesch, M. (2017). Multinationals' profit response to tax differentials: Effect size and shifting channels. *Canadian Journal of Economics*, 50(4):965–994.

Hines, J. R. and Rice, E. M. (1994). Fiscal Paradise: Foreign Tax Havens and American Business. *The Quarterly Journal of Economics*, 109(1):149–182.

Hoopes, J. L., Robinson, L., and Slemrod, J. (2018). Public tax-return disclosure. *Journal of Accounting and Economics*, 66(1):142–162.

Hosono, K., Hotei, M., and Miyakawa, D. (2018). Tax Avoidance by Capital Reduction: Evidence from Corporate Tax Reform in Japan. *Working Paper*.

Hungerman, D. M. (2013). Substitution and stigma: Evidence on religious markets from the catholic sex abuse scandal. *American Economic Journal: Economic Policy*, 5(3):227–253.

Ilchmann, C., Rösel, F., and Steinbrecher, J. (2015). Steuerwettbewerb im Kleinen Ein Blick auf den Fall Monheim. *ifo Dresden berichtet*, 22(4):26–38.

Kühn, S. (2015). *Church tax, church disaffiliation, and voluntary giving*. PhD thesis, Technische Universität Dresden, Dresden.

Langenmayr, D. and Simmler, M. (2021). Firm mobility and jurisdictions' tax rate choices: Evidence from immobile firm entry. *Journal of Public Economics*, 204:104530.

Liu, L., Lockwood, B., Almunia, M., and Tam, E. H. (2019). VAT Notches, Voluntary Registration, and Bunching: Theory and UK Evidence. *Working Paper*.

Maydew, E. L. (2001). Empirical tax research in accounting: A discussion. *Journal of Accounting and Economics*, 31(1-3):389–403.

OECD (2015). Tax Administration 2015: Comparative Information on OECD and Other Advanced and Emerging Economies. Paris.

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.

Piazza, A. and Jourdan, J. (2018). When the dust settles: The consequences of scandals for organizational competition. *Academy of Management Journal*, 61(1):165–190.

Riegel, U., Gutmann, D., Peters, F., and Faix, T. (2019). Does church tax matter?: The influence of church tax on leaving the church. *International Journal of Practical Theology*, 23(2):168–187.

Scholes, M. S., Wilson, P. G., and Wolfson, M. A. (1990). Tax Planning, Regulatory Capital Planning, and Financial Reporting Strategy for Commercial Banks. *The Review of Financial Studies*, 3(4):625–650.

Shackelford, D. A. and Shevlin, T. (2001). Empirical tax research in accounting. *Journal of Accounting and Economics*, 31(1-3):321–387.

Shevlin, T. (2020). An Overview of Academic Tax Accounting Research Drawing on U.S. Multinational Taxation. *Journal of International Accounting Research*, 19(3):9–17.

Slemrod, J. (2008). Why Is Elvis on Burkina Faso Postage Stamps? Cross-Country Evidence on the Commercialization of State Sovereignty. *Journal of Empirical Legal Studies*, 5(4):683–712.

Tennant, S. N. and 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.

Tørsløv, T. R., Wier, L. S., and Zucman, G. (2018). The missing profits of nations. *Working Paper*.

von Schwerin, A. and Buettner, T. (2016). Constrained Tax Competition: Empirical Effects of the Minimum Tax Rate on the Tax Rate Distribution. *Working Paper*.

Essay 1

Threshold-Dependent Tax Enforcement and the Size Distribution of Firms: Evidence from Germany

Threshold-Dependent Tax Enforcement and the Size Distribution of Firms: Evidence from Germany

Drahomir Klimsa*, Robert Ullmann†

Version Date: March 23, 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

*University of Augsburg.

†University of Augsburg.

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 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 Section 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 Section 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% to 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).

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

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 U.S. facing a higher IRS audit probability undertaking less aggressive tax positions compared to those facing lower audit probabilities.

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

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).

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, 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; Søgaard, 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

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.

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).

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 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

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).

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).

[Insert Table 1 about here]

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.

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.

[Insert Table 2 about here]

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 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 segmenta-

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.

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

tion 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 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 increases 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).

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

[Insert Table 3 about here]

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 Section 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 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 Section 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

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

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.

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 then 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]

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.

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.

[Insert Figure 1 about here]

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 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 Section 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 Figure 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 Section 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]

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).

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:

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

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 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_ks to

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 downwards 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 Section 2.1). However, we repeat our analyses at different placebo FSTs to ensure that our results are not driven by the selection of FSTs (Section 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.

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_ks 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_ks 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_ks 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 prefiling returns (prefilings usually occur monthly or quarterly). Table 4 shows the sample selection process.

[Insert Table 4 about here]

The data originally include 2,756,463 firms with information on both REV and EBT.[25] We first

21 Negative and significant γ_ks 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 (Section 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.

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 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 prefiling 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 $rawREV$.

[Insert Table 5 about here]

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

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 prefiling 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).

[Insert Figure 2 and Figure 3 about here]

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 Figure 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.

[Insert Table 6 and Table 7 about here]

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-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.

[Insert Figure 4 and Figure 5 about here]

Panel A of Figures 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 conditional firm audit costs. Our results correspond to the results found by Tennant 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., 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 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 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.

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% and 47.5% (OECD, 2021b), and the standard VAT rate was 19% and 21% (OECD, 2020), respectively.

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.

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

[Insert Figure 6 about here]

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

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

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 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 Figures 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 Figures 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

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.

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 Section 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 replicate the baseline analyses from Panel C of Figures 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 Section 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 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

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 a unmanaged group is not available in our setting because even FSTs for different AICs are relatively close to each other.

threshold-dependent enforcement regimes have on optimization costs.

References

Agostini, C. A., Engel, E., Repetto, A., and 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., and McKee, M. (2009). Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics*, 93(3-4):392–402.

Almunia, M. and 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. and Peichl, A. (2017). Responses of Firms to Tax, Administrative and Accounting Rules: Evidence from Armenia. *Working Paper*.

Asatryan, Z., Peichl, A., Schwab, T., and Voget, J. (2018). Inverse December Fever. *Working Paper*.

Bachas, P., Fattal Jaef, R. N., and Jensen, A. (2019). Size-dependent tax enforcement and compliance: Global evidence and aggregate implications. *Journal of Development Economics*, 140:203–222.

Bachas, P. and 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: Betriebsprüfung stärken].

Bernard, D., Burgstahler, D., and 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., and 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. and 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., and 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. and Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics*, 24(1):99–126.

Chen, C. and Lai, S. (2012). Financial constraint and tax aggressiveness. *Working Paper*.

Chen, S., Chen, X., Cheng, Q., and 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., and 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., and 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., and Shevlin, T. (2016). Financial Constraints and Cash Tax Savings. *The Accounting Review*, 91(3):859–881.

Fujii, E. T. and Hawley, C. B. (1988). On the Accuracy of Tax Perceptions. *The Review of Economics and Statistics*, 70(2):344–347.

Gallemore, J. and Jacob, M. (2020). Corporate Tax Enforcement Externalities and the Banking Sector. *Journal of Accounting Research*, 58(5):1117–1159.

Gallemore, J. and 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., and 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., and Shroff, N. (2017). Tax Rates and Corporate Decision-making. *The Review of Financial Studies*, 30(9):3128–3175.

Guedhami, O. and Pittman, J. (2008). The importance of IRS monitoring to debt pricing in private firms . *Journal of Financial Economics*, 90(1):38–58.

Gupta, S. and 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., and Rauhanen, T. (2016). The Effects of Size-Based Regulation on Small Firms: Evidence from VAT Threshold. *Working Paper.*

Harle, G. and Olles, U. (2017). *Modern Tax Audits [Die moderne Betriebsprüfung].* NWB, Herne, 3rd edition.

Henselmann, K. and 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., and Pittman, J. A. (2012). Do IRS Audits Deter Corporate Tax Avoidance? *The Accounting Review*, 87(5):1603–1639.

Hoopes, J. L., Reck, D. H., and 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., and Slemrod, J. (2018). Public tax-return disclosure. *Journal of Accounting and Economics*, 66(1):142–162.

Hosono, K., Hotei, M., and 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].* Boorberg, Stuttgart.

Kanbur, R. and Keen, M. (2014). Thresholds, informality, and partitions of compliance. *International Tax and Public Finance*, 21(4):536–559.

Klein, F. and Rüsken, R., editors (2020). *German Tax Regulation [AO § 194 Rn. 20-22].* 15th edition.

Kleven, H. J. and 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. and 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. and 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., and 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]*. Erich Schmidt, Bielefeld.

Tennant, S. N. and 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.

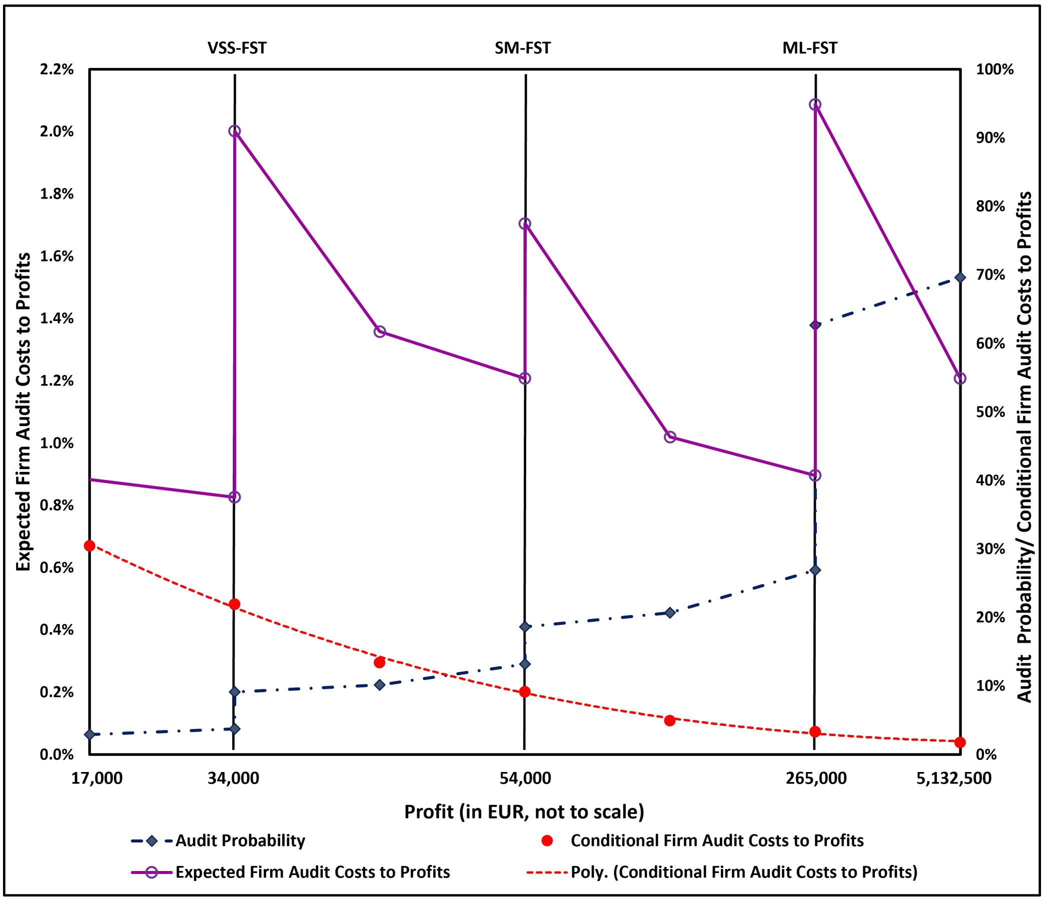
Ullmann, R. and Watrin, C. (2017). Detecting Target-Driven Earnings Management Based on the Distribution of Digits. *Journal of Business Finance & Accounting*, 44(1-2):63–93.

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.

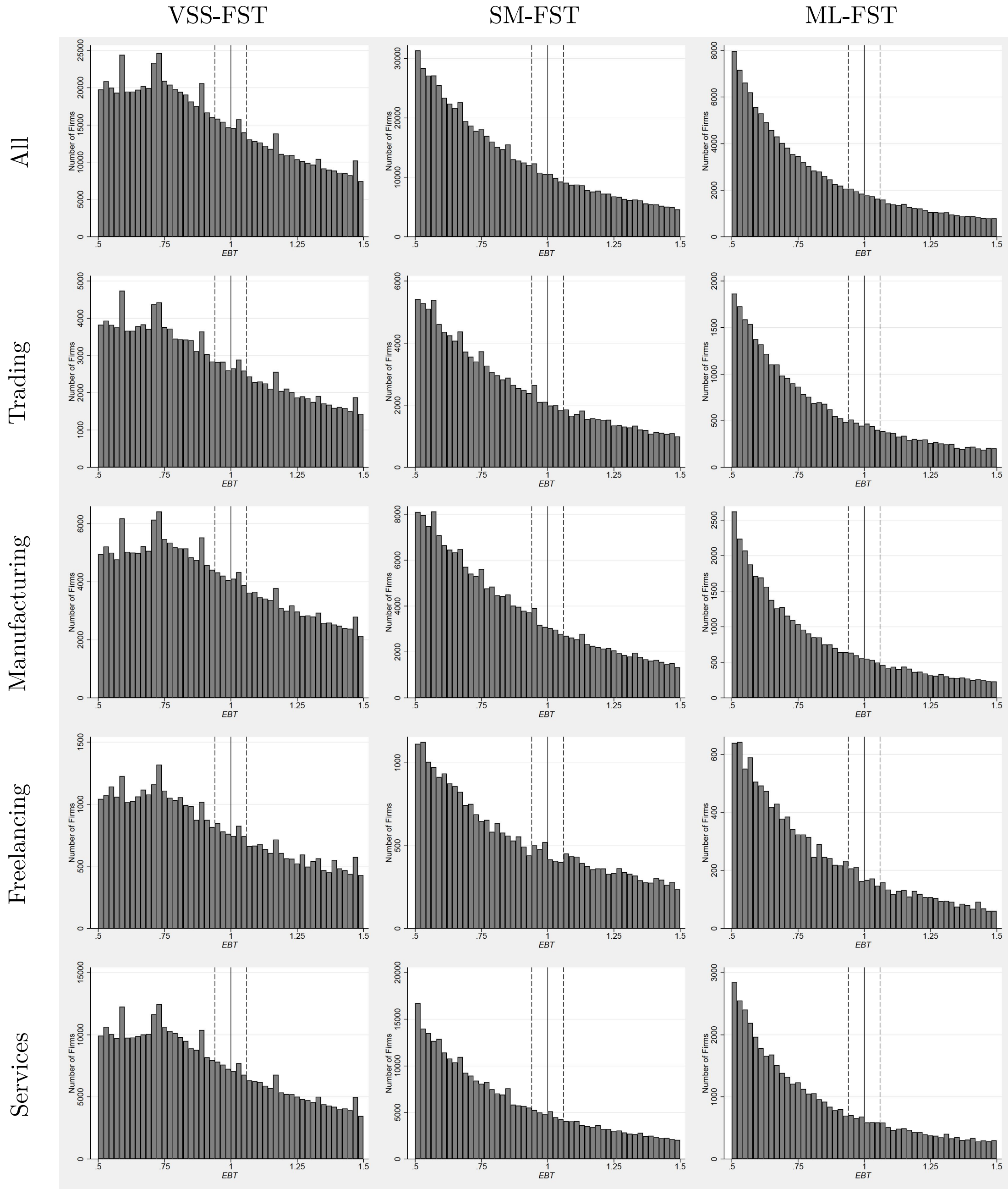
Zimmerman, J. L. (1983). Taxes and firm size. *Journal of Accounting and Economics*, 5:119–149.

Figure 1: Benefits of Size Management at Firm Size Thresholds



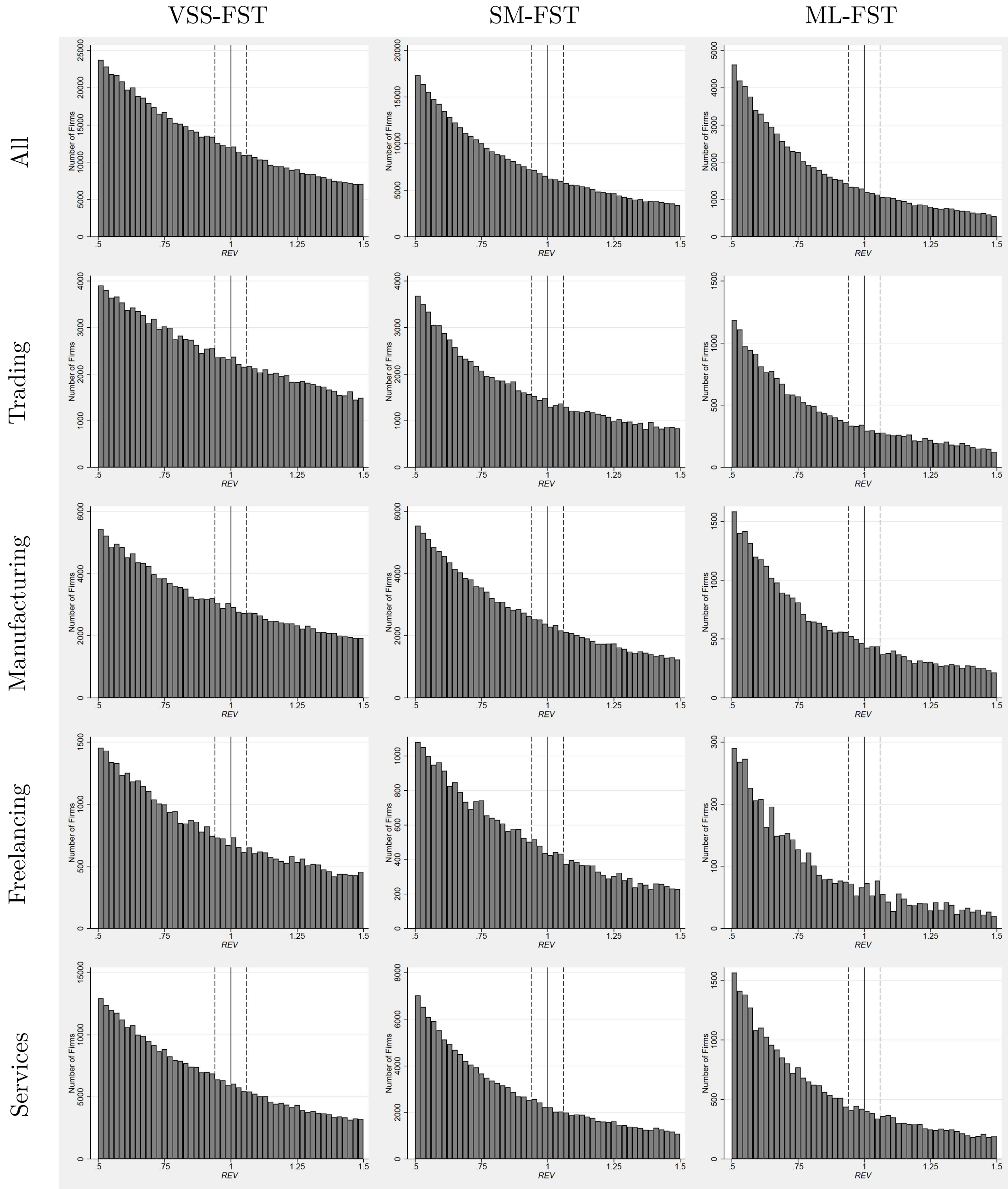
Notes: This figure graphs the ratio of conditional firm audit costs, i.e., firm audit costs once a firm is audited, to profits (dot markers) for the average firms in the VS-class, S-class, M-class and L-class and for firms at the VSS-FST, SM-FST and ML-FST (vertical solid lines) for the segmentation cycle starting in 2010 obtained from the back-of-the-envelope calculation described in Section 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 then 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.

Figure 2: Distribution of *EBT*



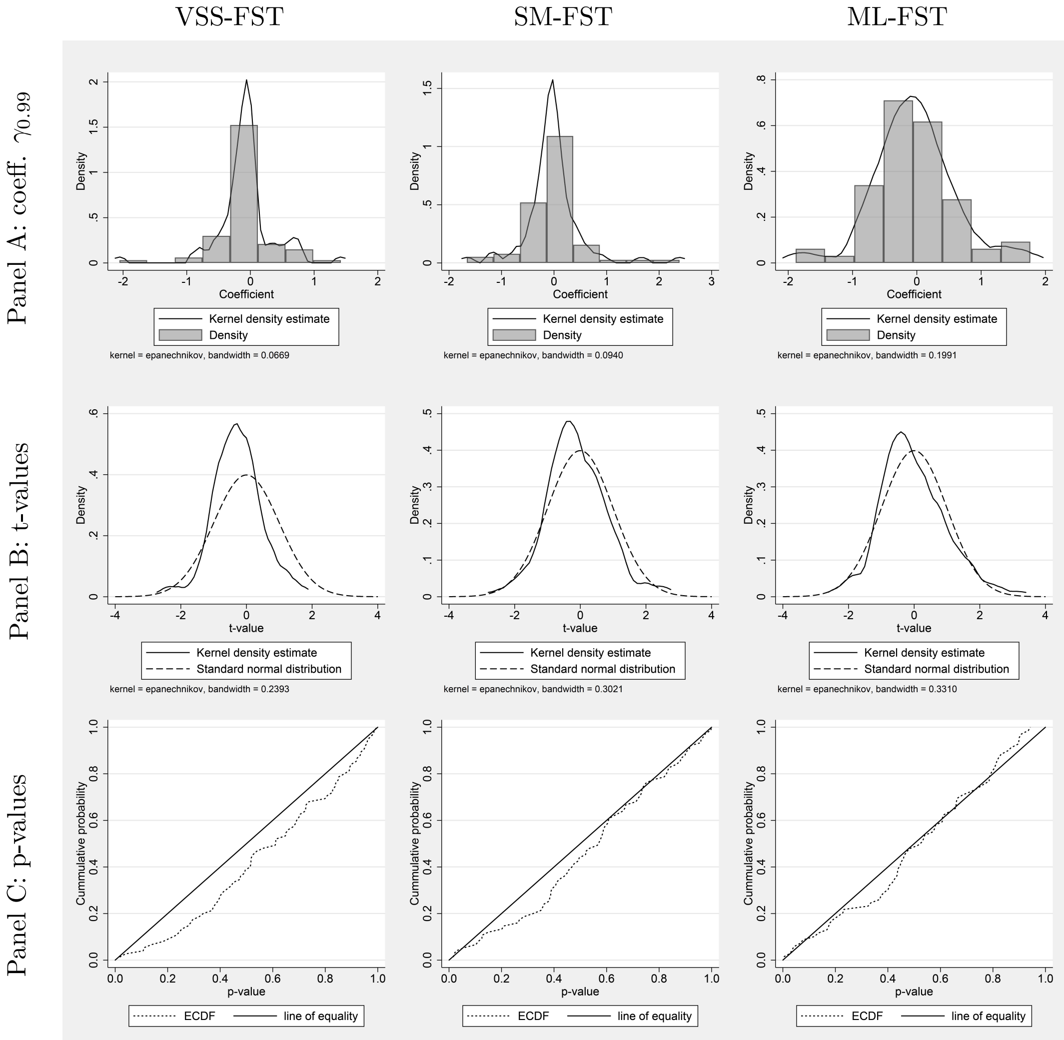
Notes: This figure graphs the distribution of *EBT* for the full sample of firms (all) and per AIC (trading, manufacturing, freelancing, services). We focus on firms in the interval [0.5, 1.5] around the VSS-FST, SM-FST and ML-FST for the segmentation cycle starting in 2010, such that the FSTs (solid 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).

Figure 3: Distribution of *REV*



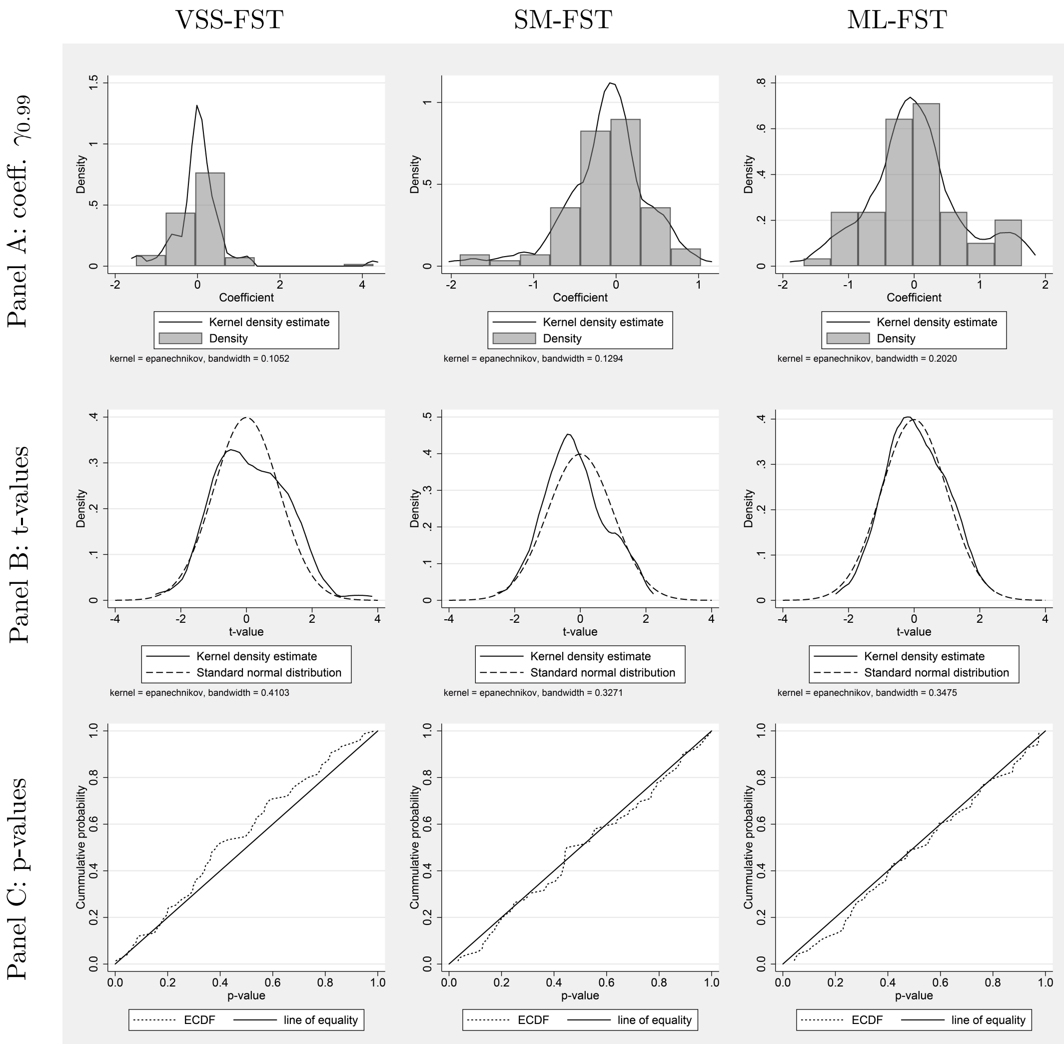
Notes: This figure graphs the distribution of *REV* for the full sample of firms (all) and per AIC (trading, manufacturing, freelancing, services). We focus on firms in the interval [0.5, 1.5] around the VSS-FST, SM-FST and ML-FST for the segmentation cycle starting in 2010, such that the FSTs (solid 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).

Figure 4: (Results for H3 (per industry): *EBT*



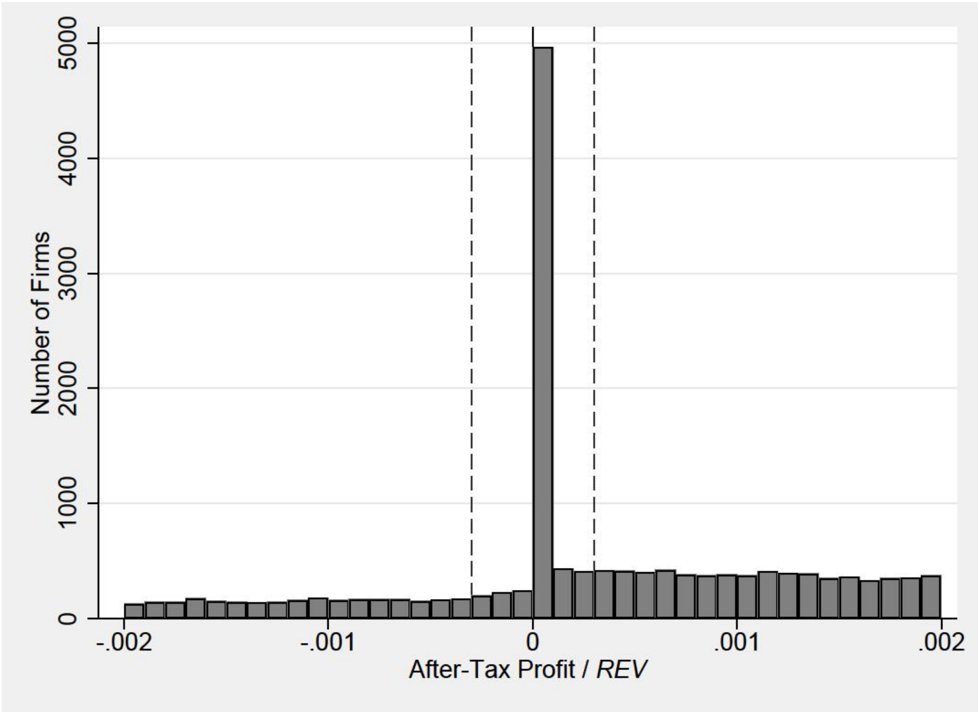
Notes: This figure graphs the distributions of regression coefficients ($\gamma_{0.99}$), t-values and p-values from estimating Equation 1 at the VSS-FST, SM-FST and ML-FST for *EBT* for the segmentation cycle starting in 2010. Panel A: Histograms and kernel density estimates (solid line) for regression 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.

Figure 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 Equation 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.

Figure 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).

Table 1: Firm Size Thresholds 2004, 2007, and 2010

AIC	Criterion	FST (in EUR)								
		Panel A: Year 2004			Panel B: Year 2007			Panel C: Year 2010		
		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

Notes: 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	2,000	3,000	6,000	2,000	3,000	15,000
		(6.7%)	(6.4%)	(2.5%)	(6.3%)	(6.0%)	(6.0%)
	Revenue	10,000	40,000	250,000	5,000	40,000	400,000
		(6.9%)	(5.3%)	(4.0%)	(3.2%)	(5.0%)	(6.2%)
Manufacturing	Profit	2,000	3,000	5,000	2,000	3,000	15,000
		(6.7%)	(6.4%)	(2.3%)	(6.3%)	(6.0%)	(6.8%)
	Revenue	10,000	20,000	200,000	5,000	30,000	300,000
		(6.9%)	(4.7%)	(5.7%)	(3.2%)	(6.7%)	(8.1%)
Freelancing	Profit	2,000	4,000	15,000	2,000	8,000	40,000
		(6.7%)	(3.6%)	(3.1%)	(6.3%)	(7.0%)	(8.0%)
	Revenue	10,000	35,000	200,000	5,000	55,000	400,000
		(6.9%)	(5.0%)	(5.4%)	(3.2%)	(7.5%)	(10.3%)
Services	Profit	2,000	4,000	15,000	2,000	4,000	25,000
		(6.7%)	(7.8%)	(5.7%)	(6.3%)	(7.3%)	(8.9%)
	Revenue	10,000	30,000	200,000	5,000	50,000	400,000
		(6.9%)	(4.8%)	(4.3%)	(3.2%)	(7.6%)	(8.2%)

Notes: 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 3: Audit Outcomes 2004, 2007, and 2010

Size class	**Firms**		**Audit rate**	**Audit period**	**Additional tax revenue**		
	N	%	**%**	**years**	**EUR total (in millions)**	**%**	**EUR per firm (in thousands)**
	Panel A: Year 2004						
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
	Panel B: Year 2007						
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
	Panel C: Year 2010						
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

Notes: 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).

Table 4: Sample Selection

Initial sample (firms with information on both *REV* and *EBT*)	**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 *EBT* and *REV*	- 678	0.02%
Final sample	**2,679,877**	**97.22%**

Table 5: Descriptive Statistics

SIZE	**N**	**%**	**Mean**	**SD**	**Q1**	**Median**	**Q3**
rawEBT	2,679,877	100.00	60,240	1,574,183	413	16,224	42,649
from CIT returns	623,221	23.26	62,218	1,564,167	-3,535	4,161	30,652
from PIT returns	497,327	18.56	144,568	1,901,604	161	19,744	82,174
from LBT returns	1,559,329	58.19	32,555	1,457,607	2,929	20,213	40,000
raw REV	2,679,877	100.00	941,742	1.93e+07	30,553	101,390	337,355
from VAT returns	2,672,639	99.73	941,361	1,93e+07	30,492	101,311	337,037
from VAT prefiling returns	7,238	0.27	1,082,396	6,992,035	49,207	132,396	486,533

Notes: 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 $rawREV$.

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

Panel A: Full sample						
FST	**Left**			**Right**		
	$\gamma_{0.95}$	$\gamma_{0.97}$	$\gamma_{0.99}$	$\gamma_{1.01}$	$\gamma_{1.03}$	$\gamma_{01.05}$
VSS	-0.1080	-0.0915	-0.1020	-0.0960	0.1490	-0.0145
	(-0.59)	(-0.50)	(-0.56)	(-0.52)	(0.81)	(-0.08)
SM	0.1050	-0.0655	-0.0264	0.0246	-0.0158	-0.0456
	(1.26)	(-0.78)	(-0.31)	(0.29)	(-0.19)	(-0.54)
ML	0.0116	-0.0069	-0.0138	-0.0089	0.0240	0.0015
	(-0.26)	(-0.15)	(-0.31)	(-0.20)	(0.53)	(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	-0.0712	-0.0021	-0.0557	-0.0268	0.0945
	(-0.63)	(-0.39)	(-0.02)	(-0.49)	(-0.38)	(1.36)
Manufacturing	-0.0931	-0.0479	-0.0682	-0.0437	-0.0291	0.0206
	(-0.52)	(-0.27)	(-0.65)	(-0.41)	(-0.31)	(0.22)
Freelancing	-0.0879	-0.1080	0.2390**	-0.116	-0.1380	-0.0342
	(-0.51)	(-0.62)	(2.43)	(-1.18)	(-0.98)	(-0.24)
Services	-0.1030	-0.1300	-0.0338	0.1130	0.0416	-0.1000
	(-0.53)	(-0.67)	(-0.26)	(0.86)	(0.50)	(-1.20)

Notes: 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).

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

Panel A: Full sample						
FST	**Left**			**Right**		
	$\gamma_{0.95}$	$\gamma_{0.97}$	$\gamma_{0.99}$	$\gamma_{1.01}$	$\gamma_{1.03}$	$\gamma_{01.05}$
VSS	-0.0238	-0.0118	0.0217	0.0082	-0.0161	-0.0447
	(-0.68)	(-0.34)	(0.62)	(0.24)	(-0.46)	(-1.28)
SM	0.0286	0.0099	-0.0192	-0.0552**	-0.0215	-0.0101
	(1.37)	(0.48)	(-0.92)	(-2.64)	(-1.03)	(-0.49)
ML	-0.0332	0.0170	0.0401	-0.0234	0.0071	0.0076
	(-0.72)	(0.37)	(0.87)	(-0.51)	(0.15)	(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	0.0161	0.0490	-0.1450**	0.0813	-0.0891
	(0.36)	(0.33)	(0.85)	(-2.52)	(0.97)	(-1.07)
Manufacturing	0.1120**	0.0008	-0.0049	-0.0342	0.0354	-0.0357
	(2.21)	(0.01)	(-0.16)	(-1.10)	(0.39)	(-0.40)
Freelancing	-0.0700	0.0901	-0.1020	-0.0891	0.158	0.378
	(-1.02)	(1.31)	(-1.15)	(-1.01)	(0.69)	(1.65)
Services	-0.0091	-0.0008	-0.0582	-0.0159	-0.0064	-0.0262
	(-0.21)	(-0.02)	(-1.48)	(-0.40)	(-0.07)	(-0.29)

Notes: 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).

Acknowledgments

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.

Data Source

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).

Threshold-Dependent Tax Enforcement and the Size Distribution of Firms: Evidence from Germany

Online Appendix A. Type II Error

Online Appendix A.1. Overview

Naturally, in an empirical test, the null cannot be confirmed. However, an investigation of the probability of type II error is needed. Specifically, a small probability of type II error increases confidence that the null is indeed true and hence that firms do not react to FSTs by size management. Our main argument is that the probability of type II error is likely small in our setting because our dataset is very large. Furthermore, as we rely on administrative data on the entire population of German firms, we also expect negligible measurement error and no selection bias. Nevertheless, below, we provide additional empirical evidence that our results are not caused by type II error.

Online Appendix A.2. Sample Size

The sample size affects the probability of type II error. Hence, we apply conventional power and sample size calculations to our baseline analyses. The regression model specified in Equation 1 is based on a fifth-order polynomial with six dummies for bins around the FST. This translates into a regression model with 11 predictors in total. For such a model, the a priori sample size calculation based on a two-tailed t-test shows that to identify a moderate effect (Cohen's f^2 of 0.15) at a significance level of 10% and with a probability of type II error not higher than 1%, i.e., a power of at least 99%, a sample size of 107 observations is required. Naturally, the probability of type II error is not higher than 1% when testing H1 and H2, where we explicitly test on the full sample. However, the sample size might go below the required minimum sample size when testing H3, where we test on AIC and industry subsamples. We report the number of observations in terms of *REV* in the interval between 50% and 150% of the VSS-FST (Panel A), the SM-FST (Panel B) and the ML-FST (Panel C) in Table A.1.

Table A.1: Distribution of Industries

NACE	**N**	**NACE**	**N**	**NACE**	**N**	**NACE**	**N**	**NACE**	**N**	**NACE**	**N**	**NACE**	**N**
Panel A: VSS-Threshold (Total = 649,380)													
47	91,757	71	13,675	32	5,851	28	3,133	33	1,847	36	727	30	367
43	86,388	55	11,992	77	5,394	31	3,083	64	1,649	24	671	08	335
56	56,517	25	9,753	74	5,095	90	2,747	27	1,274	29	659	17	322
68	50,372	93	8,554	66	4,860	95	2,724	78	1,199	14	653	91	281
96	46,340	73	8,224	52	4,385	26	2,362	02	1,140	88	584	37	221
46	33,338	10	7,627	16	4,121	01	2,259	13	1,133	20	545	87	135
45	23,621	69	7,566	79	3,977	94	2,246	80	1,127	75	531	21	128
81	21,634	82	7,261	85	3,619	92	2,235	72	1,059	50	511	51	128
49	17,295	63	6,422	86	3,301	53	2,195	22	1,040	61	456	39	123
70	17,087	41	6,282	18	3,285	58	2,064	42	1,015	11	442	03	74
62	15,538	35	6,266	23	3,158	59	1,969	38	1,003	15	429		
Panel B: SM-Threshold (Total = 383,588)													
43	63,843	71	8,769	73	3,934	66	1,853	59	1,148	36	743	15	318
47	54,779	81	7,273	16	3,634	92	1,788	38	1,124	29	696	17	313
46	25,261	41	6,964	93	3,317	58	1,552	64	934	80	648	61	229
68	23,917	55	6,827	77	3,153	78	1,447	94	933	02	563	37	197
45	17,274	70	6,551	63	3,140	74	1,426	53	929	20	550	91	154
56	15,556	62	6,437	18	2,918	85	1,396	95	928	75	535	21	129
10	10,499	32	5,604	82	2,845	27	1,358	13	818	11	438	87	123
49	10,297	86	5,161	23	2,632	79	1,300	50	795	14	419	39	89
96	10,240	35	5,121	31	2,507	22	1,283	72	768	08	418	51	63
25	9,657	52	4,280	26	2,481	42	1,233	24	758	30	368	60	54
69	9,206	28	3,955	01	2,401	33	1,227	90	751	88	359		
Panel C: ML-Threshold (Total = 82,049)													
47	9,561	71	1,643	18	857	50	637	08	343	94	229	95	90
46	9,507	62	1,633	81	847	63	598	13	310	88	177	02	77
43	7,002	70	1,594	32	833	58	591	36	308	80	175	15	68
68	6,368	26	1,266	56	821	82	563	33	303	74	174	61	63
45	3,639	69	1,208	27	773	66	515	17	300	14	172	75	57
25	3,524	96	1,196	77	756	01	467	59	295	30	152		
28	2,556	22	998	16	752	20	459	93	280	87	143		
41	2,467	35	931	55	702	64	386	85	265	92	138		
52	2,029	23	918	73	688	24	374	11	264	90	132		
10	1,690	86	901	38	656	31	371	72	244	53	116		
49	1,681	78	873	42	653	29	347	79	238	21	105		

Notes: This table reports the number of nonmissing observations per industry (2-digit NACE) in terms of *REV* in the interval [0.5, 1.5] around the VSS-FST (Panel A), SM-FST (Panel B) and ML-FST (Panel C) for the segmentation cycle starting in 2010.

For *REV*, the required minimum sample size is exceeded for the vast majority of industries. Specifically, there are only one out of 76 industries (1.32%) with fewer than 107 observations around the VSS-FST, three out of 76 (3.95%) around the SM-FST and six out of 71 (8.45%) around the ML-FST.[1] The minimum sample size is exceeded at all FSTs for all AICs. The results for *EBT* (not tabulated) show similar patterns.

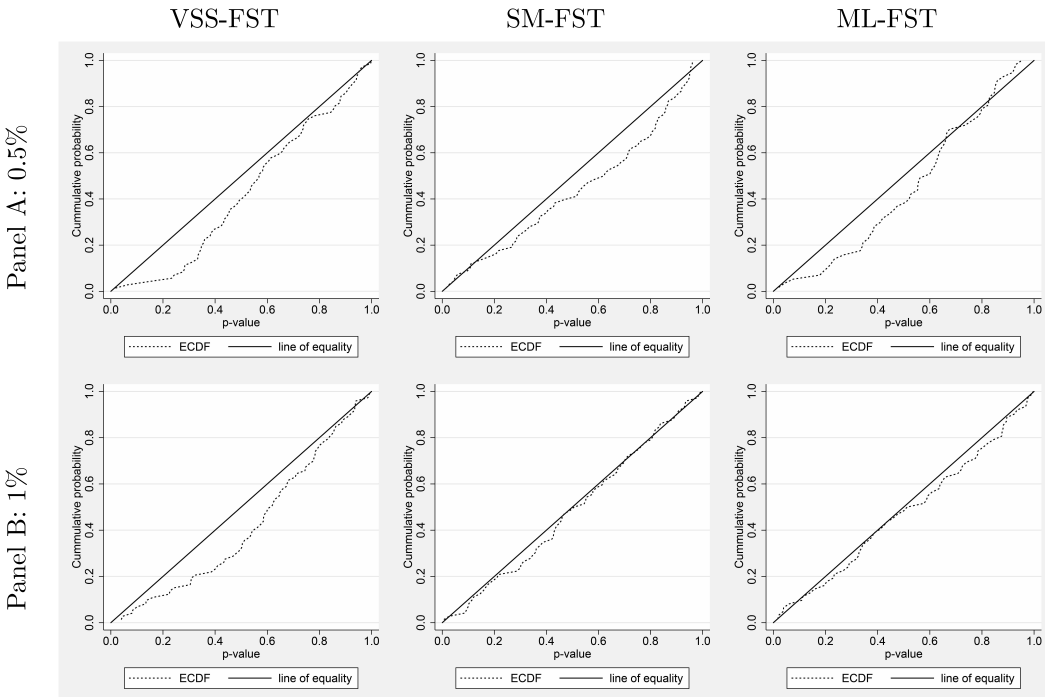
Online Appendix A.3. Parameter Specification

Furthermore, an incorrectly specified model increases the probability of type II error. Hence, we repeat the baseline analyses from Tables 6 and 7 (not tabulated) as well as Panel C of Figures 4 and 5 using two different specifications of the bin width, i.e., 1% and 0.5% of the FST. However, the results remain virtually unchanged. We report the ECDFs for the p-values per industry for *EBT* in Figure A.1 and for *REV* in Figure A.2.[2]

[1] To identify a weak effect (Cohen's f^2 of 0.02) at a significance level of 10% and with a probability of type II error not higher than 1%, a sample size of 540 observations is required. This sample size is also exceeded for the majority of industries at the VSS-FST (80.26%) and the SM-FST (78.95%) and more than half of the industries at the ML-FST (52.11%).

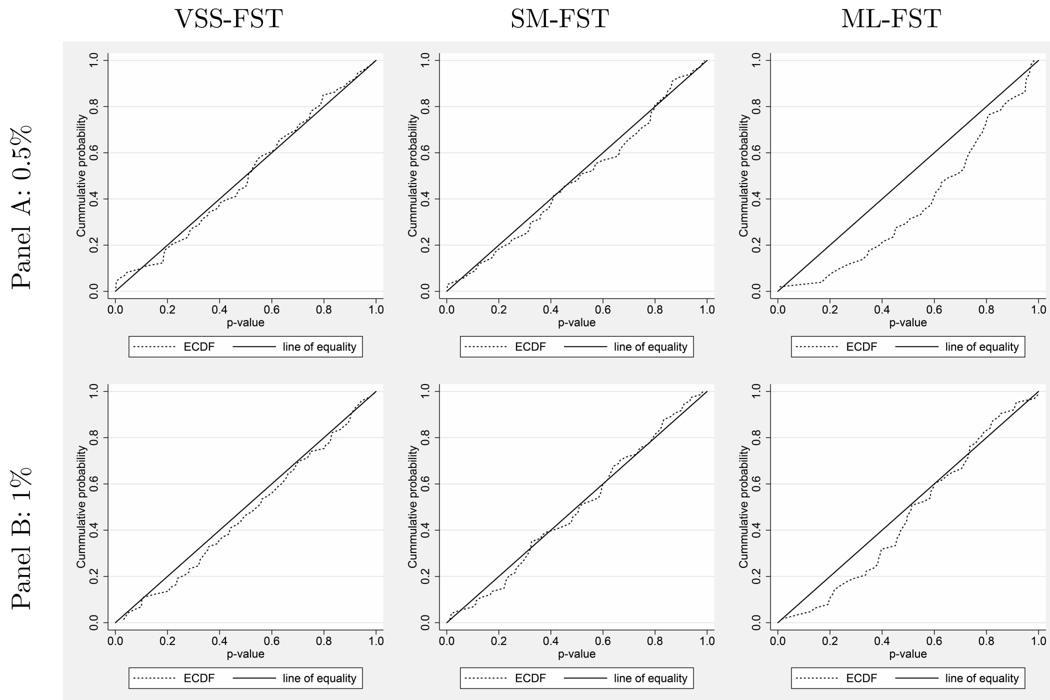
[2] Recall that the distribution of p-values provides information about size management in the overall population because under the null of an absence of size management p-values are uniformly distributed between zero and one.

Figure A.1: Distribution of p-values (per industry) for different bin widths: *EBT*



Notes: This figure graphs the distributions of p-values from estimating Equation 1 at the VSS-FST, SM-FST and ML-FST for *EBT*. Panel A: the bin width is 0.5% of the FST; Panel B: the bin width is 1% of the FST.

Figure A.2: Distribution of p-values (per industry) for different bin widths: *REV*



Notes: This figure graphs the distributions of p-values from estimating Equation 1 at the VSS-FST, SM-FST and ML-FST for *REV*. Panel A: the bin width is 0.5% of the FST; Panel B: the bin width is 1% of the FST.

Online Appendix A.4. Tests for Size Management

Finally, the choice of the statistical test also affects the probability of type II error. Hence, we repeat the baseline analyses from Tables 6 and 7 (not tabulated) as well as Panel C of Figures 4 and 5 using four alternative statistical tests to identify discontinuities in the distribution of variables (not tabulated or graphed).

As our first additional test, we use the Kleven and Waseem (2013) bunching test. This test relies on constructing a (hypothetical) counterfactual distribution of *EBT* and *REV*, i.e., a distribution in the absence of size management, which is then compared with the observed distribution to identify the excess mass below and the missing mass above the FST. Technically, the counterfactual distribution is constructed by dividing the values of *EBT* and *REV* into equal-sized bins and fitting a fifth-order polynomial using the bins as data points, excluding the bunching interval around the FST. We set the lower bound and the upper bound of the bunching interval identical to our baseline analyses. We also choose the bin width to be 2% of the FST. The bunching estimator is defined as the ratio of excess bunching over the average height of the counterfactual density in the excluded interval above the FST. Standard errors are calculated using a bootstrap procedure in which a large number of size distributions (and corresponding bunching estimates) are generated by random resampling of residuals. Consistent with Almunia and Lopez-Rodriguez (2018), we perform 200 iterations to obtain standard errors.

As our second additional test, we use the standardized difference statistic developed by Burgstahler and Dichev (1997). This method was originally applied in the context of earnings management but is applicable to our setting as well. The standardized difference approach compares the empirical number of observations in the bunching interval with the number expected in the absence of earnings management. We compute the left-sided standardized difference test statistic, which is defined as the difference between the observed number of observations in the histogram bin immediately below the FST and the average number of observations in the adjacent bin to the right and to the left scaled by its approximate standard deviation. As with our baseline analyses, we choose the bin width to be 2% of the FST.

As our third additional test, we use the nonparametric standardized difference test proposed by Lahr (2014). In this method, a kernel density distribution that is globally indistinguishable from the underlying empirical distribution of EBT, and REV is estimated and serves as a counterfactual reference distribution for the local tests. Instead of setting the optimal bandwidth a priori, the optimal bandwidth is determined by means of bootstrap simulations. As shown by Lahr (2014), the approach yields structurally equivalent but generally more conservative results compared to the approach by Burgstahler and Dichev (1997). We use an Epanechnikov kernel to estimate the counterfactual distribution.

As our fourth additional test, we implement the two-dimensional bunching test by Cox et al. (2021).[3] This test relies on constructing a two-dimensional counterfactual joint density distribution of EBT and REV, which is then compared with the observed two-dimensional joint density distribution to identify the excess mass below and the missing mass above the FSTs. Technically, the values of EBT and REV are divided into a two-dimensional grid, with grid points defined by

[3]Because firms in our setting have to simultaneously manage profit and revenue, this leads to a two-dimensional bunching problem. However, in our baseline test (and the three prior additional tests), we implicitly treat it as a one-dimensional problem by separately considering the individual dimensions, i.e., EBT and REV, because we argue that size management would create a discontinuity in the firm size distribution in the respective individual dimensions.

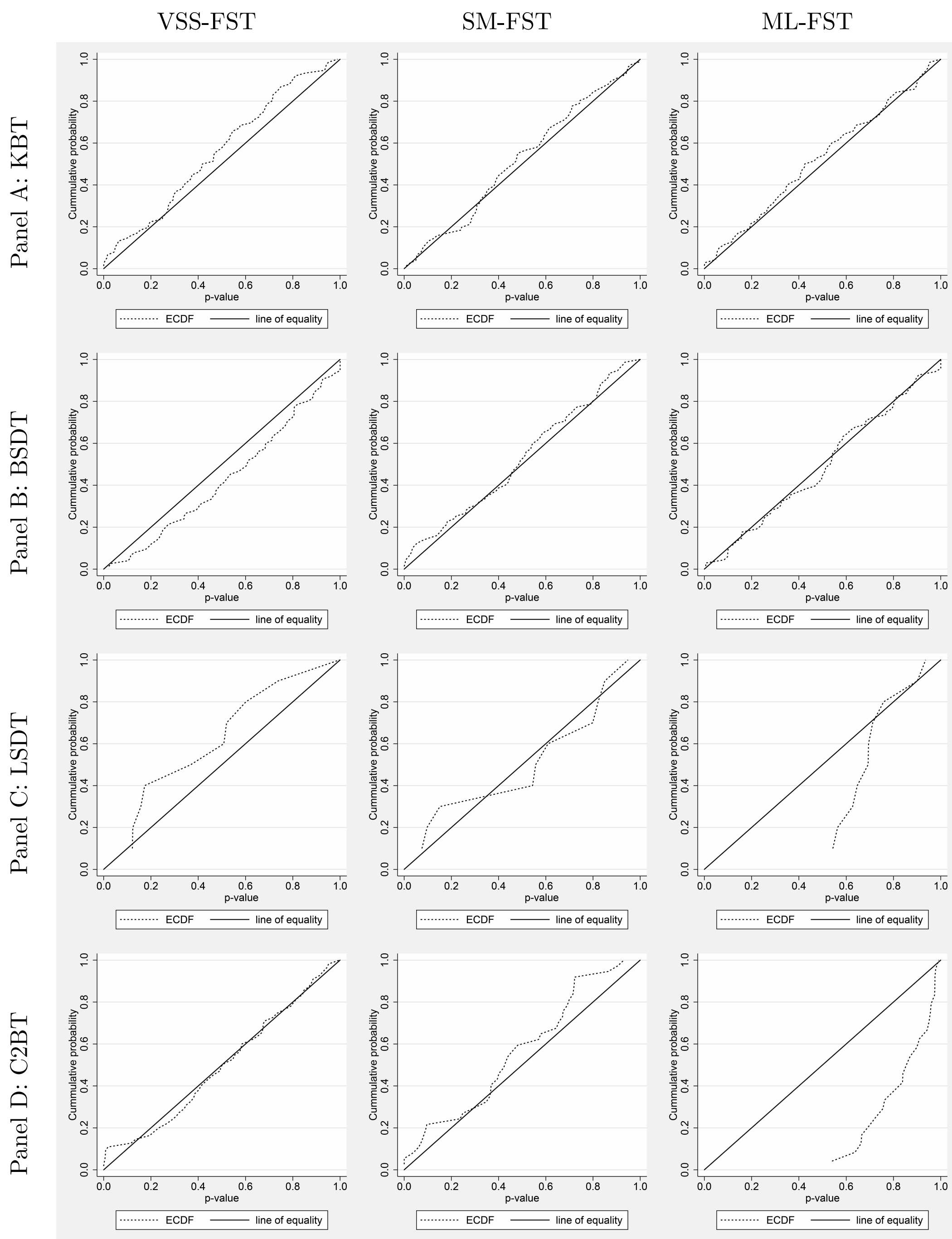
equal-sized histogram bins. The counterfactual joint density distribution is computed using nonlinear least squares and excluding the grid points in the bunching intervals around the FSTs for EBT and REV. Finally, the difference between the counterfactual joint density and the empirical joint density is computed to identify two-dimensional bunching.[4]

However, we do not find any evidence of size management around FSTs with either of these tests. We report the ECDFs for the p-values for EBT in Figure A.3 and for REV in Figure A.4.[5]

[4]We again set the lower bound and the upper bound of the bunching intervals and the bin width identical to those in our baseline analyses.

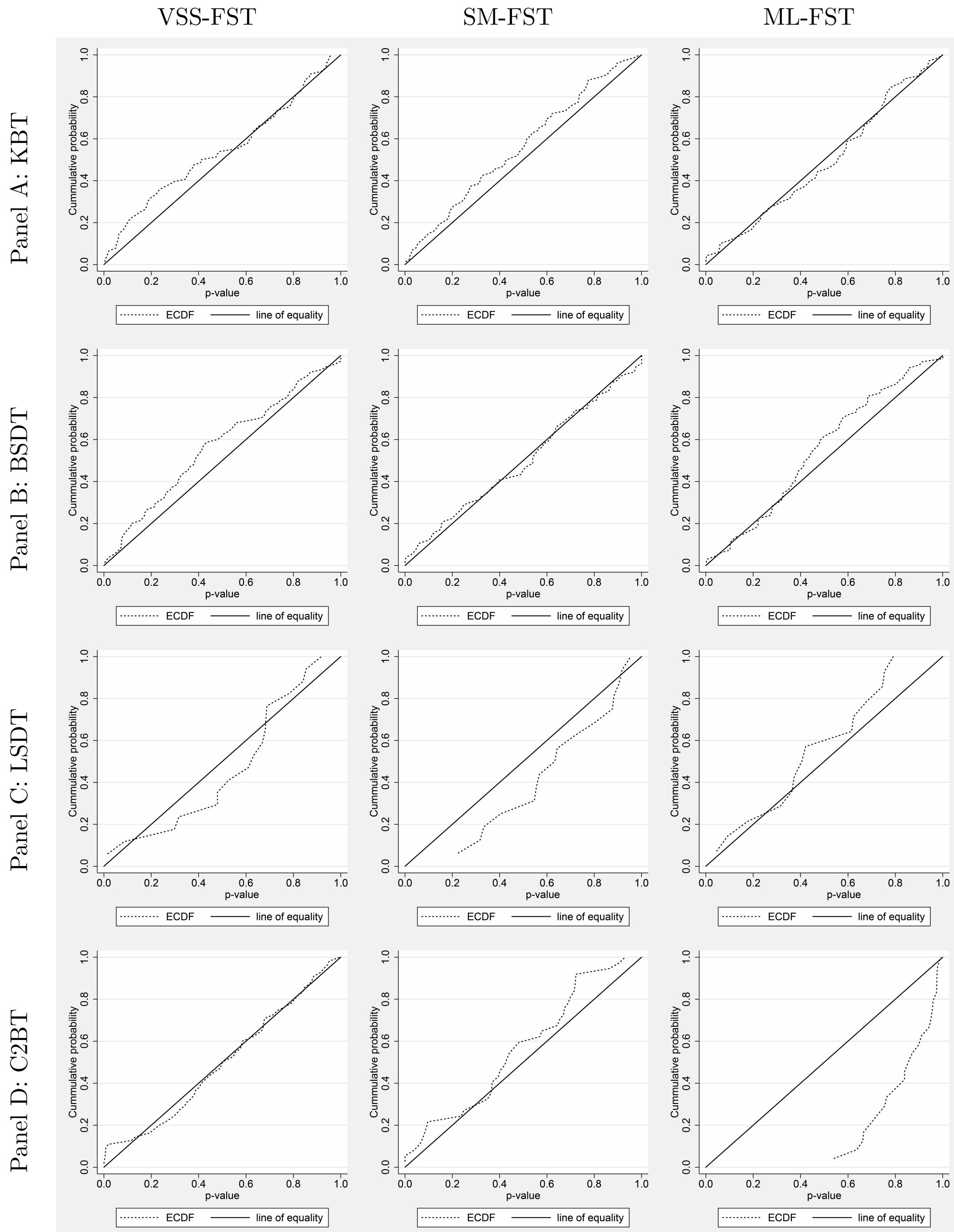
[5]Because the test by Cox et al. (2021) is two-dimensional, the ECDFs for EBT and REV reported in Panel D of Figures A.3 and A.4 are identical.

Figure A.3: Distribution of p-values (per industry) for alternative tests: *EBT*



Notes: This figure graphs the distributions of p-values from alternative tests for size management at FSTs for *EBT* (and *REV* (Panel D)). Panel A: Kleven and Waseem (2013) test (KBT); Panel B: Burgstahler and Dichev (1997) test (BSDT); Panel C: Lahr (2014) test (LSDT); Panel D: Cox et al. (2021) test (C2BT).

Figure A.4: Distribution of p-values (per industry) for alternative tests: *REV*



Notes: This figure graphs the distributions of p-values from alternative tests for size management at FSTs for *REV* (and *EBT* (Panel D)). Panel A: Kleven and Waseem (2013) test (KBT); Panel B: Burgstahler and Dichev (1997) test (BSDT); Panel C: Lahr (2014) test (LSDT); Panel D: Cox et al. (2021) test (C2BT).

References

Almunia, M. and Lopez-Rodriguez, D. (2018). Under the Radar: The Effects of Monitoring Firms on Tax Compliance. *American Economic Journal: Economic Policy*, 10(1):1–38.

Burgstahler, D. and Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics*, 24(1):99–126.

Cox, N., Liu, E., and Morrison, D. (2021). Market Power in Small Business Lending: A Two-Dimensional Bunching Approach. *Working Paper*.

Kleven, H. J. and 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.

Lahr, H. (2014). An Improved Test for Earnings Management Using Kernel Density Estimation. *European Accounting Review*, 23(4):559–591.

Essay 2

Norderfriedrichskoog! Tax Havens, Tax Competition and the Introduction of a Minimum Tax Rate

Norderfriedrichskoog! Tax Havens, Tax Competition and the Introduction of a Minimum Tax Rate

William C. Boning*, Drahomir Klimsa†, Joel Slemrod‡, Robert Ullmann§

Version Date: May 27, 2022.

Abstract

German municipalities levy local business taxes (Gewerbesteuer) by choosing a tax rate to apply to local reported business income, where the tax base is defined uniformly at the national level. Prior to the federal government's imposition of a minimum tax rate in 2004, some municipalities, such as the tiny North Sea town of Norderfriedrichskoog, chose to act as tax havens by setting a zero tax rate. We combine administrative microdata from firm tax returns with municipality-level information to study the choice of becoming a tax haven; the extent to which havens attracted income from other municipalities before and after the introduction of the minimum tax rate; and how the introduction of the minimum tax rate affected the tax competition equilibrium among non-haven municipalities. Our results suggest that income was shifted to haven municipalities both before and after the introduction of the minimum tax rate. Our findings also indicate that the mandated increase in havens' tax rates did not lead to rate increases (or decreases) among municipalities in general or tax-haven municipalities' geographical neighbors. In contrast to the literature on global business tax competition, our preferred specifications, which leverage the minimum tax rate imposition for identification, find no evidence of competition in business tax rates. We find that tax havens largely do not affect the business tax rates set by non-havens, suggesting that a global minimum tax rate binding only for international tax havens will have little effect on tax competition between non-haven countries.

Keywords: income shifting, minimum tax, tax competition, tax havens

JEL Classification: H26, H32, H71

*University of Michigan.
†University of Augsburg.
‡University of Michigan.
§University of Augsburg.

1. Introduction

International tax havens, which have long been a key component of tax planning by multinational corporations (MNCs), have drawn increased attention from policy makers and academics in recent years. In 2017, the OECD's base erosion and profit shifting (BEPS) initiative emphasized taxing profits where a firm has its presence. Although the OECD did not recommend a minimum tax rate at the time, an expert committee of the European Commission (Ruding, 1992) had advocated a legally binding minimum corporation tax rate (of 30%) almost 30years ago to prevent excessive tax competition. In July 2021, 130 countries joined the proposal of the OECD/G20 Inclusive Framework on BEPS that reallocates taxing rights on profits of large MNCs to market jurisdictions under Pillar One and establishes a global minimum tax rate of at least 15% under Pillar Two.

Extensive literature, reviewed below, is available on international tax havens. In this study, we examine a similar but distinct setting featuring tax havens in German municipalities, which can set an individual tax rate for all businesses conducted within their borders. Prior to 2004, some German municipalities functioned as domestic tax havens, levying zero or very low tax rates to attract taxable income and boost economic activity. Concerned about the revenue resulting from such practices, the German federal government drastically reduced the attractiveness of tax havens by imposing a minimum tax rate in 2004.

Studying tax havens in the German context has several distinct advantages. Germany has a single, federally-administered tax system with a fixed definition of the business income tax base. Differences in the quality of governance, which is critical to the success of international tax havens (Dharmapala and Hines, 2009; Slemrod, 2008), are not significant within Germany. Nor is the fact that the institutional features of some international tax havens are suited to particular home countries, such as the Isle of Man for British MNCs, Cayman Islands for American MNCs, and so on. Within Germany, MNCs cannot exploit differences across locations in definitions of organizational form. Moreover, the tax rate is completely transparent. Finally, administrative data from firm tax returns is available and comparable across jurisdictions. Perhaps most importantly, the effects of the establishment of a minimum tax rate of 9.1% in 2004 provide insight into the role played by tax rates in profit-shifting decisions and whether profit shifting intensifies or abates tax competition.

Of course, because of the differences in the institutional settings, we will have to be careful in extrapolating our conclusions to the international setting.

Our research addresses five empirical questions:

1. What kind of municipalities turned into tax havens?
2. What was the volume of economic activity they attracted, and how much of the tax base was shifted to these tax havens?
3. What was the nature of economic activities drawn to the tax havens?
4. What were the effects of imposing a minimum tax rate on tax havens?
5. What were the effects of imposing a minimum tax rate on non-havens?

We find that German municipal havens resemble the "dot" havens described by Dharmapala and Hines (2009)—although they are landlocked and relatively poor compared with non-havens—with minimal physical investment but substantial relocation of paper profits, both before and after the minimum tax rate was imposed.

The reminder of the paper proceeds as follows. Section 2 summarizes the related literature, Section 3 discusses the relevant German tax institutions, and Section 4 describes the data used for our analyses. Section 5 reports the results of our research and discusses the implications of these findings for policy debates about international tax havens. The final section concludes this paper.

2. Related Literature

This study connects with several studies that examine the effects of tax systems on firms' investment, location, and profit-reporting behavior; interdependence between governments when setting their potentially multidimensional tax policies; corporate tax havens; and how the interaction between firms' choices and governments' choices generates an equilibrium. Rather than attempting to review all these studies, we select a few to emphasize the link between this study and existing work on the theory and empirical measurement of the effects of tax havens and minimum tax rates.

2.1. Related Theoretical Literature

The theoretical literature on tax competition does not speak with one voice on whether tax havens lead to a reduction in tax rates elsewhere, or how a minimum tax rate would alter tax rates in jurisdictions where the tax rate lies above the minimum rate. Nor is there a consensus on the welfare effects of tax havens and minimum tax rates, although the models where tax havens reduce welfare often also predict that tax havens lead other jurisdictions to reduce tax rates. This pattern is consistent with models that assume a benevolent social planner, where competition depresses rates below the optimum but is at odds with models where tax competition constrains a Leviathan government that would otherwise set tax rates above the social optimum.

The canonical model of tax competition (Zodrow and Mieszkowski, 1986; Wilson, 1986) features competition for real activity but not shifted profits among symmetric governments. This model, in its simplest form, predicts strategic complementarity in tax rates, so that when one municipality raises rates, others follow suit.

The notion of tax havens immediately suggests asymmetry among jurisdictions, even a qualitative differentiation. Allowing asymmetric jurisdictions or a Stackelberg leader rather than simultaneous rate-setting can alter the canonical model's predictions. For example, Wang (1999) shows that imposing a minimum tax rate by preventing other municipalities from undercutting the Stackelberg leader's choice can induce the leader to reduce its tax rate. Slemrod and Wilson (2009) emphasize the role of country size in the decision to become a tax haven: the cost of forgone revenue is small relative to the potential gains from attracting income from other countries.

A comparatively smaller body of literature explicitly considers how profit shifting affects tax competition, offering diverse predictions about the effect of a minimum tax rate on tax rates of jurisdictions not directly affected, and the welfare effects of the minimum rate. Kanbur and Keen (1993) provide a model of profit shifting and the resulting tax competition, taking firm location decisions as fixed. They show that the jurisdiction with a smaller production tax base sets a lower tax rate. Imposing a minimum tax rate between the two tax rates increases both the rates in equilibrium, as well as welfare in both places.

Hong and Smart (2010) propose that tax havens raise equilibrium tax rates by enabling countries to discriminate between the mobile portion of the corporate tax base, which allows firms to use havens to lower the effective tax rate, and the immobile portion. When there are tax havens, countries can raise tax rates on relatively immobile capital without losing the mobile portion of the tax base. To the extent that raising the minimum tax rate eliminates tax havens, it would now reduce tax rates.

Johannesen (2010) adds that there is a congestion cost of an MNC shifting too much profit into a single jurisdiction, hence firms spread shifted profits across jurisdictions. Tax havens attract shifted profits that would otherwise arrive in low-tax non-haven jurisdictions. Absent shifted profits, jurisdictions have reduced incentives to set low tax rates. Eliminating tax havens by raising the minimum tax rate then provides opportunities for non-haven countries to attract shifted profits by reducing their tax rates close to the new minimum rate, and can even reduce some countries' revenues.

Including tax enforcement in a model of tax competition can change the effects of a minimum tax rate. Slemrod and Wilson (2009) propose a model where havens exert downward pressure on non-haven countries' tax rates and cause welfare losses due to the additional enforcement resources countries expend to limit the use of havens to shift profits. According to Cremer and Gahvari (1997, 2000), the enforcement policy can be relaxed to attract shifted profits, and tax harmonization without enforcement harmonization can cause intensified competition for shifted profits.

That tax rate differentials affect both the location of real activity and where taxable profits are reported is central to explaining tax havens. This is modeled in Hines and Rice (1994) by introducing a convex cost of moving a fraction of profits from local production either into or out of a jurisdiction. Grubert and Slemrod (1998) adopt a similar assumption and develop a structural model of the decision of U.S. firms to locate profits and production in Puerto Rico, given a tax provision that reduces the cost of shifting profits into Puerto Rico by locating real operating capital there. In both these models, the private cost of shifting income depends on the deviation of reported profit from real activity in the low-tax jurisdiction; therefore earning more real income in a jurisdiction reduces the cost of shifting a given amount of income, thus providing an implicit subsidy to conduct some real activity in the tax haven.

2.2. Related Empirical Literature

One strand of empirical literature focuses on describing tax havens and their activities. Dharmapala and Hines (2009) show that international tax havens tend to be small, well-governed, and often coastal or island states. Slemrod (2008) adds that they tend to also participate in other kinds of "commercialization of state sovereignty," such as money laundering and postage stamp issuance pandering, in part because they have a relatively meager endowment of standard resources.

Another strand of literature estimates the responsiveness of profit and production location to tax differentials. Grubert and Slemrod (1998) perform simulations based on empirical analysis that suggests that most U.S. firms operating in Puerto Rico would not do so in the absence of tax incentives. Hines and Rice (1994) study macro-level data and find that locations of foreign production and profits of US multinationals were highly responsive to tax rates. Dharmapala (2014) surveys the recent empirical evidence on this question. A meta-analysis by Heckemeyer and Overesch (2017) shows that averaging recent estimates using more reliable firm-level data, the semi-elasticity of reported income with respect to the cross-country difference in tax rates is 0.8, that is, reducing the difference in tax rates by ten percentage points would reduce income reported in tax havens by 8%. Studies using newer methods, however, remain divided on the role of tax havens. Dharmapala and Riedel (2013) study the pass-through effects of macroeconomic shocks in foreign non-haven countries on profits in havens and estimate that 2% of the shock to the parent's income is shifted to low-tax affiliates. Country-level descriptive statistics from the Bureau of Economic Analysis (BEA) reported in Dharmapala (2014), and international statistics, including those of tax havens in Tørsløv et al. (2018), suggest that US multinationals report around 40% of their profits in low-tax affiliates. Dharmapala (2014), however, emphasizes that these reported profits need not be reflective of the taxable income, with an estimated share of 14.5% of value added in havens.

Brueckner (2006) summarizes the empirical research on tax competition and stresses the need for the instrumental variables method to obtain consistent estimates of responsiveness to other jurisdictions' rates in the spatial lag models common in this literature. Lyytikäinen (2012) demonstrates, in the context of property taxes, that using changes in the minimum tax rate as an instrument leads to smaller estimates of tax competition than using neighbors' demographics as instruments. Besley and Case (1995) estimate spatial lag models for a variety of taxes, including U.S. state corporate income taxes.

In a much smaller body of literature, these issues have been explored in the German municipality context. Fossen and Steiner (2018) estimate that a 1% increase in the local business tax (LBT) rate decreases the LBT base by 0.45%. However, they argue that the German federal fiscal equalization scheme largely compensates municipalities for any loss in the LBT base when they increase the LBT rate and claim that the common practice of using tax budget data instead of tax return data results in a significant bias of the elasticity away from zero due to the higher volatility of tax revenues, which are largely based on prepayments in comparison to assessed taxes. Specifically, when tax rates are raised, to reduce their prepayments, firms might exaggerate the reduction in their expected tax base for the coming year when reporting to the tax. We use tax return data; therefore, our results are not subject to this concern. A study closely related to our research, von von Schwerin and Buettner

(2016), examines the effect of raising the minimum tax rate on the tax rates of geographically neighboring municipalities and of low-tax municipalities (i.e., "neighboring" in tax rate) over several years, and find that the minimum tax rate raised rates in municipalities bordering tax havens in one of these two senses. On the contrary, studying the effects of a substantial tax rate reduction in the German municipality of Monheim, Ilchmann et al. (2015) find that tax rate reductions in Monheim did not trigger tax cuts in the respective federal state. However, the findings suggest that the probability of stable or decreasing tax rates increases with the spatial proximity of a municipality. In a recent study, Langenmayr and Simmler (2021) show that municipalities set a higher tax rate if immobile firms, such as wind power plants, constitute a larger share of the municipality's tax base.

Devereux et al. (2008) provide an empirical analysis of international tax competition for both capital location and profit. They estimate reaction functions in a two-dimensional policy space where countries compete by setting both the effective marginal tax rate on investment and the statutory tax rate. The variation in tax rates they use is largely driven by rate changes in non-haven countries. They find evidence of competition in both rates, with a 1-percentage point fall in the weighted average statutory rate in other countries corresponding to a 0.7-percentage point reduction in the home country tax rate. The magnitude of the interaction between the two rates (i.e., between competition for capital and profits) is very small.

This study addresses the theoretical question of whether havens cause non-havens to raise or lower their business tax rates. The empirical literature is unsettled on the level of business income, especially the harder-to-measure business tax base, located in international tax havens. Our study estimates the level of the business tax base in havens where the tax base is more clearly defined, and the data are more comprehensive than international data. Estimates of the responsiveness of business income located in international tax havens to tax rate differentials show small effects, and evidence suggests that countries engage in substantial competition in business tax rates. In both cases, countries' endogenous choices of tax rates pose a challenge. This study contributes to both strands of this empirical literature by using variation in tax rates exogenously imposed by a minimum tax rate.

3. German Local Business Tax and its Havens

3.1. Institutions

3.1.1. Overview

Most German firms are liable to the LBT levied by municipalities.[1] In addition, federal taxes are levied, either corporate tax or income tax, are levied depending on legal form. The computation of the tax base is largely similar across these three types of taxes. Tax laws for all taxes are uniformly defined at the national level, with no differences across municipalities. Variation across municipalities in the LBT regime derives solely from differences in LBT rates. During the period 2001 to 2006, the federal portion of the overall tax rate was about half of the total tax burden, with LBT rates ranging from 0% to 15.79% at the 90th percentile and 18.03% at the 99th percentile, with a maximum outlier of 31.03%.

Non-tax laws (including labor laws) are also set at the national level and, therefore, do not vary across municipalities. Moreover, federal courts guarantee a homogeneous application of the law. General institutional factors, such as the stability of the government (central and local), functionality of public authorities, infrastructure, availability of finance, unemployment support, and antitrust regulations, are also very similar throughout Germany. Finally, because of Germany's shared culture and relatively small geographical dimensions (886 km from north to south and 636 km from east to west), any unobservable variables, such as tax-paying mentality, are arguably more uniform than in cross-border studies (Klassen and Laplante, 2012) or in within-country studies in geographically larger and culturally more heterogeneous countries, such as China (Shevlin et al., 2012) or the United States Gupta and Mills (2002).

Mainly due to EU regulations, the rules for income shifting within Germany are largely similar to the rules for cross-border income shifting. Hence, income shifting between firms in a firm group within Germany is limited mainly by the arm's length principle. This includes virtually mirroring the rules in the international context, that is, arm's length regulation for transfer prices, royalties, and interest expenses. However, in contrast to worldwide relocations or even relocations within the EU, firms relocating within Germany are not liable for any exit tax. Moreover, the German "thin

1 Exceptions are limited to farming firms as well as certain professional categories including lawyers, tax consultants and medical doctors, who are not incorporated.

capitalization rule” applied only to cross-border transactions till 2002. Starting 2003, the rule applies equally for domestic and non-domestic transactions due to EU requirements. Finally, the German-controlled foreign corporation regime does not apply to merely domestic transactions. Hence, overall, available income-shifting techniques are virtually identical for within-Germany transactions and global transactions, but domestic transactions are less restrained by tax-abuse regulations.

3.1.2. Local Business Tax

The Federal Republic of Germany has 16 states, with a total of 472 districts, in 2006. Each district is subdivided into municipalities (12,685 in 2006). LBT revenue represents a major source of German municipalities’ funds. A flat “LBT collection rate” applies to nearly all income, with full exemptions for certain unincorporated businesses such as medical practitioners, engineers, architects, lawyers, and auditors. The LBT collection rate in our sample period ranges from 0% to 900% with a mean (median) of 333% (350%), which are obviously not actual tax rates[2]. Actual LBT rates are computed by multiplying a municipality’s LBT collection rate with a statutory multiplier, where the vast majority of firms are subject to the marginal statutory multiplier of 5%[3]. Unincorporated firms also have an exemption amount of €24,500. Moreover, unincorporated firms can use an LBT credit against their income taxes; however, this arguably does not impact our results because the tax credit is, interestingly, independent of the actual LBT paid. It depends only on local business profits and is then computed based on a hypothetical average tax rate (i.e., it is the same regardless of whether the income is earned in a low-tax or high-tax jurisdiction).

Every municipality sets its LBT rate independently, but may only collect taxes from permanent establishments within its borders. Consequently, tax rate competition exists among municipal governments to encourage firms to establish facilities in their municipalities (for a detailed discussion of tax rate competition among German municipalities, see Janeba and Osterloh (2013)). Figure 1

2 Technically speaking, municipalities set an LBT collection rate. However, in the following, we refer to the LBT rate only. This is computed as LBT rate = LBT collection rate / (LBT collection rate + 2,000 %). This computation correctly considers the marginal statutory multiplier of 5 % in our sample period (paragraph 11 of Section 2 of the LBT code) and the fact that the LBT is deductible from its own tax base (known as the circularity problem of the LBT). Computation of the LBT rate from the LBT collection rate was changed structurally in 2008, that is, after our sample period, to LBT rate = LBT collection rate * 0.035 and without deductibility of the trade tax from its own tax base.

3 For very small unincorporated businesses below profits of €48,000 in our sample period, the statutory multiplier ranged from 1% to 4%. This does not hold for incorporated businesses for which the multiplier is always 5% which is also the marginal multiplier for unincorporated firms larger than €48,000 in profits.

shows the spatial distributions of LBT rates for 2001 and Figure 2 for 2006, that is, at the beginning and end of our sample period.

[Insert Figure 1 about here]

[Insert Figure 2 about here]

Tax rates vary substantially from place to place and over time. Large cities, such as Munich and Cologne, generally have higher LBT rates (often higher than 17.5%) than municipalities located in rural areas. We also observe that more affluent municipalities in the south and west levy higher LBTs than municipalities in the northeast. Moreover, we observe that municipalities with larger budget deficits, such as in the former coal-mining Ruhr areas, sometimes levy higher LBTs. However, these areas also tend to have many larger cities adjacent to one another.

Until 2003, municipalities set their LBT rates without any legal constraints. During this period, a handful of municipalities set tax rates of 0%, and several others set LBT rates between 0% and 9.1% (i.e., the tax rate that would become the threshold in 2004). The evidence we present in Section 5.3 demonstrates that these municipalities functioned as tax havens, attracting shifted profits from higher-tax-rate municipalities.

In 2004, the German federal government passed a national legislation explicitly intended to shut down municipal tax havens by disallowing LBT rates below 9.1%[4]. In 2003, the government already attempted to force tax havens to abolish their practices by allocating tax-haven income to a firm's direct parent company if the parent company was also located in Germany (in an unexpected change of the law published March 16, 2003, and already relevant for taxation in 2003[5]). Given the specific design of this law change, double taxation might have occurred because the tax base was ultimately taxed in the haven and the municipality of the parent company. Hence, this legislation specifically impacted the least aggressive tax havens, that is, those that did not have a tax rate of 0% but were still below the threshold that was used to define a haven. Immediately after the law change, firms in Norderfriedrichskoog appealed to the fiscal courts regarding the 2003 tax rate change; however, the Federal Fiscal Court, which is Germany's highest in tax and fiscal matters, ruled on July 4, 2010,

4 The minimum required LBT collection rate is 200%, corresponding to a tax rate of about 9.1% = 200% / (200% + 2,000%), as discussed in footnote 2.

5 Bundesgesetzblatt 2003 Teil I Nr. 19, 20.05.2003. Note that, at that time municipalities, could only increase their trade tax collection rate until June 30th of a given year while they could decrease it until December 31st.

that the tax would have to be paid and ruling that the law was brought in existence according to formal constitutional regulations (leaving the more important question of material constitutionality unanswered). Similarly, Beiersdorf-Freudenberg, another tax haven at that time (zero tax rate in 2003, 0.01% tax rate in 2002), directly appealed to Federl constitutional Court, which is Germany's supreme court, regarding the 2004 law change and refused to collect the LBT from its businesses. Beiersdorf-Freudenberg was allowed to temporarily defer tax collections until the court ruled on January 25, 2005, that the LBT would have to be collected both for 2003 and 2004. Due to various reasons, particularly the aforementioned unintended targeting of the least aggressive tax havens, relatively straightforward strategies to circumvent the new law (i.e., by establishing foreign direct parents for tax haven subsidiaries), the retroactive introduction, and a relatively widespread belief that the entire law could be unconstitutional, the 2003 law change was abolished again on December 23, 2003, and a minimum LBT rate of 9.1% (relevant for taxation in 2004 and thereafter) was introduced.

In 2001 (2006), the beginning (end) of our period of investigation, the average LBT collection per municipality in our data was €2.03 million (€3.13 million), amounting to about 13.7% (21.3%) of average municipality income (including fees, fines, interest income, rent income, etc.). Municipalities also collect taxes on real estate, for which they can also set individual tax rates.

3.1.3. Fiscal Equalization Scheme

Germany maintains a multi-layered fiscal equalization scheme intended primarily to smoothen volatility in LBT receipts. During the study period, this scheme included income tax, value-added tax (VAT), and LBT. Income tax and VAT are first collected by the federal government and then partially redistributed to the municipalities (sometimes indirectly via redistribution among the states). Depending on the exact state redistribution scheme, municipalities receive a fraction of approximately 12-18% of the income tax collected from their residents, which seeks to compensate for the difference between municipality of residence and municipality of business. Municipalities also receive about 2% of the overall VAT revenue; however, the exact amount depends on various parameters, such as the LBT revenue of the municipality.

On the other hand, the LBT collected is to be partially forwarded to the federal government and state governments. Forwarding towards the federal and state governments follows a rule that depends on the LBT rate (with variation in parameters per year, and between states). As a general rule, the lower the LBT rate, the higher the share of LBT tax collected that must be redistributed among other municipalities (with a maximum of 100%, which, however, is only binding if the municipality strongly decreased its rate relative to previous years). Furthermore, there are redistribution schemes between state governments and municipalities, and municipalities within state; these are based on state-level legislation. Hence, they vary by state and year, but with notable similarities among states. In most states, the key parameters of reallocation are the number of inhabitants, LBT revenue, number of students in school, number of low-income families, and so on.

Overall, the fiscal equalization scheme partially mitigates tax competition between municipalities but does not eliminate it (for more details, see Fossen and Steiner (2018)).

3.2. Norderfriedrichskoog: An Illustrative Example

According to contemporaneous news reports, around the year 2000, the German North Sea village of Norderfriedrichskoog consisted of only 13 farmhouses and less than 50 inhabitants. Until 2003, it was the poster child for municipal tax havens, being at least the nominal home of subsidiaries of companies such as Deutsche Bank, Lufthansa, and power and gas giant E.ON, apparently because the local authority did not levy any LBT.

In fact, Norderfriedrichskoog was founded, in 1696, on a tax exemption. Around 300 years ago, a local duke issued a tax exemption in return for building a dike to keep the North Sea out (Oberteis, 2002). Given the low need for revenue in the absence of public facilities, the mayor set the LBT at 0% in 1978, and companies began locating to Norderfriedrichskoog under a federal tax office ruling that some aspects of communications and core management operations must be based in a municipality to claim its low tax rate. According to newspaper accounts (Schmidt, 2008), a local farmer established an office service and rented out rooms in her farmhouse—at one point there were 19 firms based there—and managed several of those firms. It was claimed that as many as 130 jobs were created in and around Norderfriedrichskoog since 1992, a substantial number considering that the population of the village was only 45 in 2003.

Norderfriedrichskoog was not the only German municipality tax haven but was by far the most important of the small number of municipalities setting such low business tax rates. Contemporaneous media reports particularly mention other zero-tax municipalities (Clorius, 2008). After the minimum tax rate was raised, Norderfriedrichskoog continued to impose the minimum tax rate threshold of 9.1.% even as companies began to leave. Eventually, however, it was forced to raise tax rates further due to a change in the state-level fiscal equalization scheme, which included a de facto minimum rate higher than the federal minimum (Clorius, 2008).

News reports about Norderfriedrichskoog's heydays echo the stories of prominent international tax havens. The reports also suggest that the minimum tax rate dramatically reduced businesses' use of Norderfriedrichskoog as a haven. These accounts serve as an inspiration for our empirical analysis that follows.

4. Data

We merge confidential administrative firm-level data from VAT returns for the years 2001 to 2006 with publicly-available municipality-level macroeconomic data from the Research Data Center (RDC) of the Federal Statistical Office and the Statistical Offices of the Federal States and with geographical data from the German Federal Agency for Cartography and Geodesy.

The VAT-return data cover the full population of German firms above the threshold of €17,500 per annum. Although firm-level LBT-return data would also be available from the German Statistical Office, the VAT-return data have three main advantages. First, the VAT-return data are available annually, while the LBT-return data are only available for research in waves every three years. Second, some firms liable for VAT are not liable for LBT, and these firms may form a useful comparison group in the analysis at some point. Third, the VAT thresholds are much lower than are the LBT thresholds, and hence we can observe firms that are very small and that would potentially not be included in the LBT-return data.

The main limitation of the VAT-return data, which would also apply to the LBT-return data, is that the firm group's tax returns cannot be linked. Thus, the unit of observation in our data is a single entity firm-year (specifically including information about the municipality location of each firm in each year). The VAT-return data include data from the company registry that particularly provides the number of employees, a field that is filled in for about 70% of observations; we add one to the listed number of employees, which could represent the owner and accounts for sole proprietorships. We also have information on sales subject to tax at the normal 16% VAT rate, the reduced 7% VAT rate, and tax-exempt sales, as well as information on input costs (excluding labor costs), with observations being nearly perfectly filled with all the data derived directly from the VAT-return forms. We do not have information on capital stocks or investments.

The German Statistical Office's municipality-level macroeconomic data include information on the number of inhabitants, surface area and land use, number and use of buildings, migration, election results, tourist visits, and others, in addition to information about tax rates and the municipality revenue structure. Where municipality level variables are not available (i.e., for certain years or certain municipalities), we try to estimate relevant variables from corresponding district-level data. Finally, we add data from the German Federal Agency for Cartography and Geodesy, which include mapping information of each municipality as well as the location of borders and geographical municipality midpoints.

We merge these data sources using municipality identifiers. Our sample includes, for example, for 2006, 12,180 out of Germany's overall 13,085 municipalities (i.e., 93.1%). However, this ratio is driven downwards by municipalities that cannot have firms (called non-municipality-related areas; usually bare mountains, lakes, and extensive wooden areas), that is, respective non-merges are indeed based on actual circumstances rather than missing data. We leave these in the raw data set of publicly available data before merging with the tax return-data, because it is impossible to clearly identify these with the publicly available data we have (it is easily possible based on the confidential data). The ratio of merges is lower in earlier years (e.g., 86.9% in 2001) as the fiscal authorities improved data quality significantly around the beginning of our sample period.

A discussion on our proxy for the LBT base per firm can be found in the appendix. Ultimately, we estimate firm-level profit based on sales and input costs which are available in our VAT-return data. We perform correlation tests of our estimated tax base and the observed LBT collected, and find a positive correlation close to one.

5. Empirical Analysis

5.1. What Kind of Municipalities Became Tax Havens?

For our descriptive analysis, we define tax havens as those municipalities that set an LBT rate of less than 9.1% in 2002, one year before the public debate leading up to the reform. We also drop those municipalities for which VAT-return data could not be matched for at least one firm in at least one year. By this definition, there are 27 tax havens, including Norderfriedrichskoog. Recall that Dharmapala and Hines (2009) noted that international tax havens are typically small (population below one million), more affluent than other countries on average, and well-governed; they tend to be islands (i.e., not landlocked), closer to financial centers, and possess lesser natural resources than non-havens. German municipal tax havens are overwhelmingly small and have (uniformly) good governance. In terms of GDP per capita and unemployment, however, tax havens within Germany tend to be poorer than other municipalities and are located mostly in the low-income eastern part of the country.

Table 1 compares German municipal tax havens (Panel A) to other German municipalities of similar population size (Panel B), that is, municipalities with a population smaller or equal to the biggest tax haven rounded up to 10,000, and all non-haven German municipalities (Panel C) during 2001-2003, that is, the pre-minimum tax period.

[Insert Table 1 about here]

The average non-haven municipality (of similar population size as tax havens) in Germany has 6,754 (2,103) inhabitants, gains 13.34 (4.71) new residents per year, and has a population density of 180.41 (124.85) inhabitants per km^2. The surface area is 27.72 (21.40) km^2. 57.90% (59.25%) of the surface area (is used for farming and 0.80% (0.56%) is used for business. The price of one m^2 of land is €73.13 (€65.12). The unemployment rate is 5.06% (5.16%) and GDP/capita is €20,259 (€19,663).

All the tax havens, except the largest (Nuthe-Urstromtal, 7,271 inhabitants in 2002) have fewer inhabitants than the national average of non-havens, with a mean (median)[6] of 1.083 (613) inhabitants across all tax havens.[7] In contrast to non-havens, tax havens also suffer from depopulation with a mean (median) of - 3.78 (- 4.00) residents per year. With 47.62 (31.2) inhabitants per km^2 and a surface area of 36.29 (15.03) km^2 tax havens also tend to be larger in area and less densely populated than non-havens.[8] We note that consistent with the expectation stemming from the international setting, German tax-haven municipalities tend to be relatively small on average in terms of population. However, they tend to be large in terms of surface area, indicating that they tend to be located in rural areas. At 56.75% (68.14%), a similar share of the surface area is used for farming;

6 We note that the mean in the tax haven sample is particularly subject to influence by outliers because there are only 27 tax havens.

7 We also note that three of the zero-rate tax havens do not report any inhabitants in our data in 2002, that is, Sachsenwald, Buchholz and Solling. These are most likely measurement errors, but in any case, this error will be corrected when merging the municipality-level data with tax-return data, as municipalities without firms will mechanically be dropped then. We leave these three municipalities in our sample for now.

8 Nine of the 27 tax havens are larger in area than the national average of non-havens.

however, at 0.35% (0.08%) the share of the area used for business is substantially smaller. The price of 1 m^2 of land is also substantially lower than that of non-havens, amounting to only €22.12 (€21.75). Unemployment is relatively high, at 8.72% (5.56%) and GDP per capita is relatively low at €16,760 (€14,811). Hence, contrary to the international setting, tax havens within Germany also tend to be poorer than other municipalities. This finding also holds when we compare tax havens to non-havens of similar population size and located in the same states as havens (not tabulated).

Next, we analyze the spatial distribution of tax havens. Figure 3 maps the geographical location of the 27 tax havens identified in 2002 within Germany (red dots for tax rate exactly zero; blue dots for LBT rates below 9.1% but greater than zero) and the districts to which they belong (yellow areas).

[Insert Figure 3 about here]

It is apparent from this map that most of Germany's tax havens in 2002 were located in states that were historically part of East Germany, and the eastern region remains economically underdeveloped compared with the more affluent states that were historically part of West Germany. One noteworthy exception is Norderfriedrichskoog, which is the red-dot tax haven located in the far north near the Danish border on the German coast of the North Sea.

In summary, Germany's domestic tax havens exhibit only some of the characteristics attributed to international havens. While they share the nationwide good governance and rule of law and tend to be small in terms of population size, they are relatively poor and far away from Germany's economic centers in the west and south.

5.2. What Was the Volume of Economic Activity Drawn and Tax Base Shifted to Tax Havens?

While the LBT base in tax havens is large per inhabitant, tax havens make up only a small portion of the overall LBT in Germany. On average, the estimated total profits reported in tax havens each year before the minimum LBT rate was imposed range between 0.027% and 0.237% of the national LBT base in Germany (not tabulated). We repeat this investigation using only sales and find that sales by tax-haven firms are between 0.032% and 0.113% of overall sales in Germany over the sample period (not tabulated). This is much less than what was estimated by prior literature for the international setting, that is, 6% of all corporate profits and 10% of all corporate tax revenue (Tørsløv et al., 2018).

If that tax base was instead taxed at 14.01%, which is the national average of LBT rate for all municipalities before the minimum LBT rate was imposed, weighted by the estimated LBT base of each municipality per year, total revenues would be 0.05%, or €77.5 million, higher. However, as the incentive to relocate profits to tax havens is strongest in the highest-rate municipalities, profit shifting from these municipalities should be higher. Moreover, in the absence of tax havens, the average LBT rate might be higher. Consequently, using the observed weighted average national rate may understate the revenue costs of tax havens in Germany. On the other hand, higher tax rates resulting from the absence of tax havens could lead to tax avoidance by other means or to reductions in economic activity, which would lead to our estimate overstating the potential revenue gains from eliminating tax havens.

5.3. What Was the Nature of Economic Activity Drawn to Tax Havens?

We mostly use graphical analysis to investigate the nature of economic activity attracted to tax havens in terms of indicators of pure shifting rather than real activity, such as the ratio of employees per inhabitant and per firm. We relied on confidential data for this analysis. Hence, we need to ensure that confidentiality requirements are met, including the requirement that scatter plot dots can only be shown if they include at least three observations (i.e., three firms or three municipalities depending on the analysis). We construct bins based on the 2002 LBT rate, starting at the lowest LBT rate (i.e., zero). For the most aggressive tax havens, we first consider a tax rate of zero (red dots). We then loop through all the years of our sample period to ensure that we have at least three observations in the bin each year. If at least one year does not fulfill the requirements, we add the next higher 2002 LBT rate and repeat the procedure till the requirement is met. We then use the next highest 2002 LBT rate as our new starting point for the less aggressive tax havens (blue dots) and repeat the procedure. We continue to use the procedure to create bins ranked according to the 2002 LBT rate for which we can ensure the confidentiality requirement. Hence, each bin remains constant over the entire sample period, enabling us to gain intuition about the development of tax-haven municipalities relative to non-haven municipalities over time. We do this separately for each variable of interest as discussed below. Thus, dots might not comprise the exact same municipalities for each variable of interest, depending on the data availability for that variable.

Dots show the average LBT rate as well as the average for the variable of interest (computed as a simple average over the municipality-year-level ratios) over all municipalities within the bin. The size of the bubbles indicates the number of municipalities included following our stepwise procedure discussed above. The grey vertical line shows the threshold rate of 9.1% which applies since 2004. The results for employees per inhabitant (employees per firm) are shown in Figure 4 (Figure 5).

[Insert Figure 4 about here]

[Insert Figure 5 about here]

Figure 4 shows that tax havens have the largest number of employees per inhabitant of any municipality from 2001 to 2006. In some years, the tax havens contain more employees than inhabitants, that is, the ratio of employees per inhabitant is greater than one—not even considering the inhabitants who are not available to the workforce (e.g., children and the elderly). This effect is likely either due to commuters for employment into the municipality or due to inhabitants with more than one (part-time) job as unemployment rates are relatively high in tax havens (Table 1). Figure 5 shows that havens have a similar number of employees per firm as other places albeit they lie at the lowest end of the distribution. This indicates that firms in tax havens do not require many employees. Figure 6 below shows the ratio of our proxy for the LBT base at the municipality level to inhabitants. Figure 7 shows the net sales per inhabitant.

[Insert Figure 6 about here]

[Insert Figure 7 about here]

Figure 6 shows that the red-dot tax havens also host remarkable amounts of tax base per inhabitant. This does not appear to change considerably after the introduction of a minimum tax rate in 2004 , and throughout our entire sample period. Moreover, 2002 was apparently an outlier year for tax havens (or for one of the tax havens in the red dot), as the LBT base was orders of magnitude higher than in any other year. The same pattern holds for sales per inhabitant, which is shown in Figure 7—we observe that actual sales volume is shifted into the red dot tax havens in 2002. One may speculate that the 2002 outlier year might have spurred political interest in tax havens and led to the introduction of a minimum threshold in 2004. Overall, the results provide some indication that activity in tax havens represents income shifting by "paper transactions" rather than real activity, that is, physical production, as we observe a relatively low number of employees per firm and extraordinarily high LBT bases and sales per inhabitant in tax havens compared with non-havens. Figure 8 shows the actual LBT collected in €1,000 per inhabitant.

[Insert Figure 8 about here]

The impression from Figure 6 and Figure 7 is confirmed when looking at Figure 8. Not surprisingly, tax havens hardly collect any LBT until 2003, as they apply zero or a low tax rate to their tax base (tax base estimates being indicated in Figure 5). Starting 2004, havens collect by far the highest LBT per inhabitant, orders of magnitude higher than the other bins—in €100,000s rather than low single digits. Notably, the value per inhabitant stays constant after the introduction of the minimum threshold and even increases slightly between 2004 and 2006. This indicates that taxable income did not shift away from tax havens to a significant extent in the years immediately after the introduction of the minimum tax rate.[9]

Next, we compare the industry breakdown of tax havens and similar non-haven municipalities (in terms of population size as detailed above) that are located in the same states in 2002 and 2005 (Figure 9).

[Insert Figure 9 about here]

Figure 9 shows that the share of firms operating in the agriculture, forestry and fishing industry (industry 1) is relatively high in tax havens compared with similar non-havens. Furthermore, tax havens have a relatively low percentage of firms operating in producing industries (industry 7), construction (industry 10), and wholesale and retail (industry 11). This observation is consistent with the fact that tax havens are relatively poor in terms of GDP per capita and suffer from relatively high unemployment rates as their economies are mostly based on industries with low added value. In contrast, the percentage of firms operating in real estate activities, renting of moveable property and services for enterprises (industry 15) is relatively high in tax havens. This in turn is consistent with the notion that economic activity in tax havens is rather based on income shifting than real economic activity because such activities are particularly suited for inter-company paper transactions as we know from international tax havens.

9 We quantify the differences between tax havens and non-havens more precisely in regression analyses, which confirm the impressions formed by the graphs shown above (not tabulated). Tax havens tend to have significantly fewer employees per firm, as well as more firms and a higher LBT base and sales per inhabitant.

5.4. What Were the Effects of Imposing a Minimum Tax Rate on Tax Havens?

The reported sales and the LBT base in tax havens decline only slightly in the three years after the minimum tax rate was introduced, as is visible from the time series statistics in Figure 6 and Figure 7. If tax rates elsewhere did not react substantially to the minimum tax rate, these time series are consistent with a relatively inelastic response of reported profit and tax base location to changes in tax rate differentials, as the tax rates of the most desirable havens rose by 9.1 percentage points.

5.5. What Were the Effects of Imposing a Minimum Tax Rate on Non-Havens?

Tax havens can affect the tax rates that non-havens set through two channels, which we examine separately. First, firms' ability to shift profits to tax havens via paper transactions may cause non-havens' tax bases to be more or less responsive to the tax rates non-havens set, changing their revenue-maximizing LBT rates. Second, tax havens may also compete with non-havens for the location of physical production, again altering the responsiveness of non-havens' LBT bases.

The first channel, tax havens' profit-shifting effects, do not vary with the distance between havens and other municipalities as profit-shifting costs are independent of distance. The 2004 introduction of a minimum LBT rate of 9.1% enables us to study the profit-shifting effects of tax havens by inspecting the patterns of LBT rate changes in and after 2004. We test the predictions made by i) the Keen and Konrad (2013) and Slemrod and Wilson (2009) models, that is, non-haven tax rates will rise, against the prediction by ii) Hong and Smart (2010) that non-haven tax rates will fall, and iii) Johannesen (2010) prediction that non-haven tax rates may fall for some municipalities that set rates above the minimum. Both the direct inspection of LBT rate changes and spatial auto-regression with distance-independent weights rely entirely on the time series for identification, as the treatment-the 2004 change in tax rates of havens–does not vary across non-havens in these models. Directly inspecting the LBT rate changes allows us to look separately at subcategories of the rate changes, splitting increases from decreases and dividing further by the domain over which rates changed.

On the contrary, for the second channel, competition for physical production is more intense between nearer neighbors due to the distance-dependent costs of relocating production. We use spatial auto-regressive models with a variety of weights that capture functions of distance. We compare estimates of the intensity of tax competition from i) OLS models, ii) models using neighbors' demographics as instruments for the neighbor's LBT rate, iii) and models using LBT rate increases forced by the minimum tax rate as instruments. Geographical differences in treatment intensity allow us to identify the effect of the 2004 change independent of the time-series shocks that affect all municipalities similarly. Using progressively better-identified instruments, the magnitude of our estimates of tax competition decreases, until our preferred specifications, that is, using the changes forced by the minimum rate as an instrument for changes in neighbors' tax rates, estimated in differences from 2002 to 2004, finds no evidence of substantial competition in business tax rates among geographical neighbors. This is also consistent with the notion that activity in tax havens relies rather on paper transactions than physical production.

5.5.1. Paper-Transaction Profit-Shifting Channel

We compare changes in the LBT rates set by non-havens in 2004, when the minimum LBT rate threshold became formally binding, to changes in the LBT rates set by non-havens in other years. Doing so enables us to assess whether the introduction of the minimum LBT rate resulted in unusual patterns of rate changes attributable to the minimum threshold (i.e., the shutdown of tax havens). Table 2 provides details of the distribution of rate changes in each year.

We assume that the LBT rates in each year are in Nash equilibrium so that 2004 is the only "treated" year[10]. The distribution of the rate changes supports this assumption. There is substantial yearly turnover in tax rates, with approximately 10% of municipalities changing their business tax rates annually. Additionally, no municipality required to raise rates to meet the 2004 minimum rate "overshoots" the minimum threshold, in contrast to what one would expect if rate setting was governed by an s-S model of rate-setting, where there are fixed costs of adjusting the tax rate, due to political concerns for example.

[Insert Table 2 about here]

Imposing a minimum tax rate of 9.1% in 2004 appears to have had little, if any, effect on the number or extent of non-haven tax rate increases or decreases. Excluding the 20 municipalities that

10 As discussed above, a ruling that came into effect in 2003 constitutes a form of partial treatment, although the legal requirement a minimum tax rate did not take effect until 2004. The results taking 2003 as an additional "treated" year would not differ substantially, as is evident in Table 2.

increased tax rates in 2004 to exactly the required minimum, 7.91% of municipalities raised rates in 2004, which is 0.6 percentage points lower than the average across other years, but still more than in either 2002 or 2006. Again, excluding increases to the minimum threshold, the average LBT rate increase in 2004 conditional on a rate increase was 0.81 percentage points. This is 0.03 percentage points more than the average rate increase conditional on an increase in other years. Slightly fewer municipalities reduced rates in 2004 than in other years, 0.62%, which is 0.1 percentage points lower than the non-2004 average. The rate decreases conditional on a decrease in 2004 were comparable to those in other years, averaging 1.36 percentage points, 0.02 percentage points higher than the non-2004 average. While the mean change conditional on any change is 1.39 percentage points higher in 2004 than in other years, if the 20 rate increases to the minimum are included, restricting the analysis to non-havens' LBT rates, the difference is only 0.043 percentage points. Averaging after including municipalities that do not change rates, tax rates rise slightly less in 2004 than in other years, by less than 1/100th of a percentage point. The introduction of the minimum threshold of 9.1% in 2004 does not substantially alter the tax rate distribution, neither raising rates as Slemrod and Wilson (2009) would predict, nor reducing them as Hong and Smart (2010) would predict. Finally, few non-havens adjust rates down to the minimum rate in 2004, in contrast to the prediction made by Johannesen (2010). In 2004, six municipalities reduced rates to the minimum rate, roughly on a par with the number of such reductions in prior years. Non-havens do not appear to reduce rates substantially to obtain a share of the additional shifted profits associated with having a tax rate tied to the lowest rate.

While we find that the minimum threshold does not substantially alter the overall distribution of tax rate changes, suggesting that incentives for profit-shifting do not drive overall rate-setting behavior, localized competition for physical production may still lead havens' neighbors to change their rates in response to the minimum rate .

5.5.2. Production-Location Profit-Shifting Channel

We assess the nature of tax competition between neighbors for physical production before and after the minimum threshold was introduced, using spatial autoregressive models, as is common in the tax competition literature (Brueckner, 2006). We weight neighboring municipalities' tax rates by measures of distance, and we interpret the results in which competition is more intense between jurisdictions geographically closer as evidence of competition for production location.

The regression equation for the tax rate reaction function we specify is

$$\tau_{it} = \beta \sum_{j \neq i} w_{ij} \tau_{jt} + \gamma' X_{it} + \mu_{st} + \varepsilon_{it'}, \tag{1}$$

where the coefficient of interest is β, the slope of the reaction function, τ_{it} is the LBT rate of municipality i in year t, and the spatial weights w_{ij} are calculated as described below, X_{it} is a vector of municipality demographics; μ_{st} is a state-by-year fixed effect, and ε_{it} is an error term. Municipality demographics include the surface area and fractions of surface area in use overall, for settlement and streets, for mining, and for farming; the total population, and fraction of the population that is female; births per capita; deaths per capita; youth per capita; and elderly per capita. The weights capture several functions of the distance between municipalities. Letting $f(d_{ij})$ denote the function of distance (and, in some cases, population), each set of weights is

$$w_{ij} = \frac{f(d_{ij})}{\sum_j f(d_{ij})}. \tag{2}$$

Three sets of weights capture the most extreme gradients by distance: one set assigns equal weight to the three nearest neighbors by midpoint distance; another to the five nearest neighbors by midpoint distance; and a third to all municipalities in the same district, which is the smallest geographic division above municipalities. Sets of weights also include inverse distance, inverse distance squared, population-weighted inverse distance, and inverse distance squared. A final set of weights counts the population-weighted inverse distance squared only to the ten largest cities by population. Weighting by population captures the economic importance of a municipality j, and one might expect that tax rates are especially responsive to the rates of the most populous places.

OLS regressions of this form face the critique that unobserved factors may cause tax rates in jurisdiction i and in its neighbors j to move in sync for reasons other than tax competition, biasing estimates of β upward. While state-by-year fixed effects capture many such factors, they fail to capture spatially auto-correlated shocks that vary within states. State-by-year fixed effects also account for states' fiscal equalization schemes, which vary from year to year.

We compare these coefficients from OLS with two-stage least squares specifications using the demographics of neighboring municipalities as instruments for the neighbors' tax rates, as in most of the tax competition literature surveyed by Brueckner (2006) and as applied by Devereux et al. (2008) to international business tax competition. We also use increases in tax rates required to comply with the minimum tax rate as instruments for neighbors' tax rates, as Lyytikäinen (2012) does in the context of property tax competition.

Neighbors' demographics may fail to be exogenous for several reasons, as discussed in Lyytikäinen (2012). The primary concern is that any spatially correlated omitted variables that are correlated with the demographic variables themselves will also be correlated with the error term in the regression equation. Spatially correlated shocks to the demand for government funds may arise through changes in omitted demographic variables such as unobserved changes in age structure. Neighboring jurisdictions' demographic variables may also be correlated with the error term in the own tax rate specification if they make locating production in the neighbor more attractive, for example if demographic changes are correlated with the quality of the available workforce.

OLS estimates of the tax rate reaction to neighbors rates, which are likely biased upward as discussed above, are substantially above zero for nearly all weights in both the periods before and after the minimum tax rate was introduced. Taking these estimates naively, this would suggest substantial tax competition for production location in business tax rates. Table 3 presents results in detail.

[Insert Table 3 about here]

In Table 3, the estimated coefficient on the weighted average tax rate of other jurisdictions $\bar{\tau}_{jt}$ is statistically different from zero at conventional confidence levels for all the weight measures except population/distance squared to the ten largest cities and population/distance in both time periods (2001-2003 and 2004-2006). Significant coefficients are largest when weights are proportional to inverse distance or assign positive weight only to municipalities in the same district and vary greatly across weight measures from approximately 1 to around 0.25. Values before and after 2004 are generally similar within each weight measure.

Given that OLS estimates of tax competition may be biased, we instrument for neighbors' tax rates with neighbors' demographic characteristics, which reflect the need for government revenue to provide local public goods as well as the ability of a municipality to generate business income. Devereux et al. (2008) use demographic instruments to study international competition in business tax rates, which are commonly used in the empirical literature on tax competition (Brueckner, 2006). Corresponding results are still subject to concerns about the exogeneity of the demographic instruments used, driven primarily by spatial autocorrelation. The results are presented in Table 4.

[Insert Table 4 about here]

The instruments predict the weighted average tax rate strongly as the first stage F-statistics are large. The results shown in Table 4 reject the null hypothesis of a zero-reaction function slope in both periods for all weights. The point estimates are generally smaller than the OLS results, consistent with an upward bias in the OLS results. The point estimates again vary across weights, from less than 0.1 for population/distance squared and population/distance squared to only the ten largest cities to around 0.8 for 1/distance. Estimated reaction function slopes after 2004 are not consistently larger or smaller compared with those for the same weights before 2004, suggesting that the minimum threshold of the LBT rate binding tax havens does little to constrain tax competition for physical production.

These results are comparable to those of Devereux et al. (2008) when using comparable weights and statutory tax rates. While we do not weight municipalities uniformly, and cannot weight by municipal GDP or FDI flows, we can approximate the importance of size using population/distance. Our point estimates for 2001-2003 and 2004-2006 using population/distance are 0.374 and 0.439, broadly comparable to their point estimates of 0.357 using GDP weights and 0.340 using FDI weights. Their point estimate of 0.678 using uniform weights is in line with our findings of substantial tax competition with coefficients ranging from 0.1 to 0.8 using all weights except the two incorporating population/distance squared. Our specifications do differ from theirs in that we do not employ municipality indicators or time trends, while they include both for countries. However, we include state-by-year fixed effects, which remove much of the same variation their fixed effects do. They incorporate the effective tax wedge, capturing responses in the definition of the tax base, which is fixed in our context, and the R-squared of their regressions is 0.93, while ours is around 0.55, perhaps because of the additional predictive power of country fixed effects. Using demographic instruments, it appears that municipalities compete in business tax rates as much as countries do.

We use the required increase in LBT rates from 2002 to 2004, that is, the increase necessary to comply with the LBT minimum threshold, as an instrument first in regression specifications in levels, as above, and second in a regression in differences between 2002 and 2004. Both regressions produce substantially different results. The regression results in levels are shown in Table 5, while those in differences are shown in Table 6. The major difference between the specifications is that differencing removes time-invariant factors affecting the tax rate and neighbors' tax rates, including the time-invariant components of demographic variables, so that the minimum instrument becomes more influential relative to the background characteristics.

[Insert Table 5 about here]

[Insert Table 6 about here]

In the levels regressions in Table 5, some weights produce smaller point estimates than those using demographics as instruments, while others produce larger point estimates, and all reject the null hypothesis of no strategic interaction with $p<0.05$, both before and after 2004. All specifications in Table 5 reject the null hypothesis of no strategic interaction. In the differences regressions (2002-2004 differences) in Table 6 all weight definitions fail to reject the null hypothesis of strategic interaction at conventional confidence levels; for several weights, the estimates are precisely specified around zero, with confidence intervals that exclude even the point estimates using demographic instruments above and in Devereux et al. (2008). The low statistical significance in Table 6 is attributable in part to absolutely small changes in the point estimates and, in part, to the imprecision of the differences regressions for the population-weighted measures, where one would expect that the minimum-instrument specifications would be imprecise because the populations in havens are small, and for the weight based on inverse distance.

More specifically, the results reported in Table 5 imply that tax competition for production location is less intense than the results suggest in Table 4, this is, using demographics as instruments. Focusing solely on the pre-2004 period, the point estimates weighting by the closest three or five municipalities are both 0.04, versus 0.202 and 0.240 using demographic instruments. Point estimates weighting by municipalities in the same district or by inverse distance or inverse distance squared remain similar to their values using demographic instruments, while the estimates using population-weighted inverse distance and inverse distance squared rise to 1.022 and 0.853, respectively. Coefficients are roughly similar in the post-2004 and pre-2004 periods.

Investigating Table 6 in more detail, there is no statistical evidence that neighbors respond to the tax havens' change in LBT rates. The effects are precisely estimated to be zero using weights of 1/distance squared, the three or five closest municipalities, and other municipalities in the same district, ruling out substantial reactions by those municipalities most likely to be affected by the change in havens' competitiveness for real production location. The effects are estimated very imprecisely for the population-weighted measures, as one would expect given that the tiny populations of tax havens result in little variation in the population-weighted average tax rate. Results weighting by 1/distance are also imprecisely estimated, perhaps for similar reasons related to the variation in the instrument.

To summarize the empirical results on tax competition, we find that, in response to a substantial 9.1 percentage point reduction in tax rate differentials, non-havens adjust their tax rates no differently in 2004 than in other years, in contrast to the theoretical predictions of Slemrod and Wilson (2009) and Hong and Smart (2010). Consistent with Lyytikäinen (2012) results using variation from a minimum tax rate to instrument for Finnish property tax rates, we find that the estimated slope of the reaction function is lower using the minimum instrument in levels and in 2002-2004 differences instead of the demographic instruments, and both sets of instruments produce smaller point estimates than OLS. We find no statistical evidence of spatial competition in business tax rates in regressions where the outcome variable is the difference between 2002 and 2004 LBT rates.

Taken together, we find that tax havens largely do not affect the business tax rates set by non-havens, when looking at the paper-transactions channel or the real production-location channel, suggesting that a global minimum tax rate applied only to international tax havens would have little effect on tax competition between non-haven countries.

6. Conclusions

This study assesses the role played by tax havens as low-tax hosts for business profits, leveraging data on the LBT, a business income tax collected by German municipalities, with variation provided by the introduction of a minimum LBT rate of 9.1% in 2004. The minimum rate also enables us to estimate how responsive business tax rates are to exogenous changes in haven tax rates. We study the nationwide change in 2004, when the minimum rate was introduced, to understand how reducing the incentives for profit-shifting to tax havens affects non-haven tax rates. Spatial auto-regression allows us to assess whether changes in tax havens' LBT rates have especially strong effects on the rates set by their nearest neighbors, which could reflect competition for the location of physical production.

German municipal tax havens such as Norderfriedrichskoog share small populations and good governance with international "dot" tax havens but are landlocked and relatively poor compared with non-havens; they host significantly more sales, firms, and employees (likely part-time due to a management location requirement for access to their low tax rates) per capita than non-havens. However, our results indicate that activity in tax havens represents income shifting by paper transactions rather than real activity, this is, physical production, as we observe a relatively low number of employees per firm and extraordinarily high LBT bases and sales per inhabitant in tax havens compared with non-havens, and a high percentage of firms operating in real estate activities, renting of movable property, and services for enterprises.

We estimate that, in the three years before 2004, between 0.027% and 0.237% of the national LBT base, or €0.28 to €2.47 billion, was located in tax havens with rates below 9.1%. We estimate that a one-percentage point increase in haven tax rates reduces the tax base located in tax havens by about 4%.

We find that havens have no substantial effect on the tax rates of non-havens through the paper-transactions profit-shifting channel, and no significant effect on the tax rates of their neighbors through the production-location channel in our preferred specification using changes forced by the minimum tax rate to instrument for changes in neighbors' tax rates. In contrast to OLS results, results using the standard demographic instruments, and results in levels using the minimum instrument, once regressions are specified in differences from 2002-2004 using the changes forced by the minimum tax rate as instruments, we do not reject the null hypothesis that municipalities compete in LBT rates. The results are for many weights precisely estimated, with 95% confidence intervals excluding the point estimates obtained using demographic instruments in this study and in the international context by Devereux et al. (2008) and even a reaction function slope of 0.1 for the nearest neighbors. The minimum instrument is better identified than demographic instruments, given concerns about spatial autocorrelation. However, using the minimum instrument immediately raises questions about the local average treatment effect it identifies, as the minimum binds only for about 20 tax haven municipalities out of more than 12,000 municipalities overall. Our population-weighted estimates using this minimum instrument are estimated imprecisely, as one would expect given that the minimum affects only sparsely populated tax havens. We argue that tax havens host little production activity, therefore havens may not compete with their neighbors for physical production even if other jurisdictions do.

We conclude that tax havens in the German context resemble international corporate tax havens in that the tax base located in these havens is responsive to an exogenous change in the tax rate differential between havens and non-havens. Profit shifting, once established, does not seem to shrink when rate differentials shrink, pointing to short-term fixed costs of profit relocation rather than variable costs. If reductions in rate differentials do not cause large changes in the tax base reported in tax havens internationally, as within Germany, recent reductions in corporate tax rates by big countries, including the U.S., may not immediately alter the profits located in international tax havens; nor will imposing a minimum worldwide tax rate, even with a common consolidated corporate tax base. Forcing tax havens to raise their business tax rates neither raises tax rates across the board through dramatic reductions in profit-shifting nor does it lead tax havens' neighbors to raise rates. To the extent that Norderfriedrichskoog and its likes can inform us about the role of profit shifting in tax competition, the evidence suggests that profit shifting, while costly in tax revenue terms, does not substantially change the elasticity of the corporate tax base.

References

Besley, T. and Case, A. (1995). Incumbent behavior: Vote seeking, tax setting and yardstick competition. *American Economic Review*, 85(1):25–45.

Brueckner, J. K. (2006). Strategic interaction among governments. In Arnott, R. J. and McMillen Daniel P., editors, *A Companion to Urban Economics*, pages 332–347.

Clorius, S. (2008). Tax havens: Thirteen courtyards, five hundred companies.

Cremer, H. and Gahvari, F. (1997). Tax competition and tax evasion. *Nordic Journal of Political Economy*, 24(2):89.

Cremer, H. and Gahvari, F. (2000). Tax evasion, fiscal competition and economic integration. *European Economic Review*, 44(9):1633–1657.

Devereux, M. P., Lockwood, B., and Redoano, M. (2008). Do countries compete over corporate tax rates? *Journal of Public Economics*, 92(5):1210–1235.

Dharmapala, D. (2014). What do we know about base erosion and profit shifting? a review of the empirical literature. *Fiscal Studies*, 35(4):421–448.

Dharmapala, D. and Hines, J. R. (2009). Which countries become tax havens? *Journal of Public Economics*, 93(9):1058–1068.

Dharmapala, D. and Riedel, N. (2013). Earnings shocks and tax-motivated income-shifting: Evidence from european multinationals. *Journal of Public Economics*, 97:95–107.

Fossen, F. M. and Steiner, V. (2018). The tax-rate elasticity of local business profits. *German Economic Review*, 19(2):162–189.

Grubert, H. and Slemrod, J. (1998). The effect of taxes on investment and income shifting to puerto rico. *The Review of Economics and Statistics*, 80(3):365–373.

Gupta, S. and Mills, L. F. (2002). Corporate multistate tax planning: benefits of multiple jurisdictions. *Journal of Accounting and Economics*, 33(1):117–139.

Heckemeyer, J. H. and Overesch, M. (2017). Multinationals' profit response to tax differentials: Effect size and shifting channels. *Canadian Journal of Economics*, 50(4):965–994.

Hines, J. R. and Rice, E. M. (1994). Fiscal paradise: Foreign tax havens and american business*. *The Quarterly Journal of Economics*, 109(1):149–182.

Hong, Q. and Smart, M. (2010). In praise of tax havens: International tax planning and foreign direct investment. *European Economic Review*, 54(1):82–95.

Ilchmann, C., Rösel, F., and Steinbrecher, J. (2015). Steuerwettbewerb im kleinen ein blick auf den fall monheim. *ifo Dresden berichtet*, 22(4):26–38.

Janeba, E. and Osterloh, S. (2013). Tax and the city — a theory of local tax competition. *Journal of Public Economics*, 106:89–100.

Johannesen, N. (2010). Imperfect tax competition for profits, asymmetric equilibrium and beneficial tax havens. *Journal of International Economics*, 81(2):253–264.

Kanbur, R. and Keen, M. (1993). Jeux sans frontières: Tax competition and tax coordination when countries differ in size. *The American Economic Review*, 83(4):877–892.

Keen, M. and Konrad, K. A. (2013). Chapter 5 - the theory of international tax competition and coordination. In Auerbach, A. J., Chetty, R., Feldstein, M., and Saez, E., editors, *Handbook of Public Economics vol. 5*, volume 5, pages 257–328. Elsevier.

Klassen, K. J. and Laplante, S. K. (2012). Are u.s. multinational corporations becoming more aggressive income shifters? *Journal of Accounting Research*, 50(5):1245–1285.

Langenmayr, D. and Simmler, M. (2021). Firm mobility and jurisdictions' tax rate choices: Evidence from immobile firm entry. *Journal of Public Economics*, 204:104530.

Lyytikäinen, T. (2012). Tax competition among local governments: Evidence from a property tax reform in finland. *Journal of Public Economics*, 96(7):584–595.

Oberteis, R. (2002). Taxes saved behind the dike.

Ruding, O. (1992). *Report of the Committee of Independent Experts on company taxation. Executive summary. March 1992.*

Schmidt, W. (2008). Norderfriedrichskoog - the lost tax haven behind the dike.

Shevlin, T., Tang, T. Y. H., and Wilson, R. J. (2012). Domestic income shifting by chinese listed firms. *Journal of the American Taxation Association*, 34(1):1–29.

Slemrod, J. (2008). Why is elvis on burkina faso postage stamps? cross-country evidence on the commercialization of state sovereignty. *Journal of Empirical Legal Studies*, 5(4):683–712.

Slemrod, J. and Wilson, J. D. (2009). Tax competition with parasitic tax havens. *Journal of Public Economics*, 93(11):1261–1270.

Tørsløv, T. R., Wier, L. S., and Zucman, G. (2018). The missing profits of nations. *National Bureau of Economic Research*, (Working Paper no. 24701).

von Schwerin, A. and Buettner, T. (2016). Constrained tax competition empirical effects of the minimum tax rate on the tax rate distribution. In *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Fiscal Competition No. G10-V1*, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Fiscal Competition.

Wang, Y.-Q. (1999). Commodity taxes under fiscal competition: Stackelberg equilibrium and optimality. *American Economic Review*, 89(4):974–981.

Wilson, J. D. (1986). A theory of interregional tax competition. *Journal of Urban Economics*, 19(3):296–315.

Zodrow, G. R. and Mieszkowski, P. (1986). Pigou, tiebout, property taxation, and the underprovision of local public goods. *Journal of Urban of Economics*, 19(3):525–542.

Figure 1: Spatial Distributions of LBT Rates (2001)

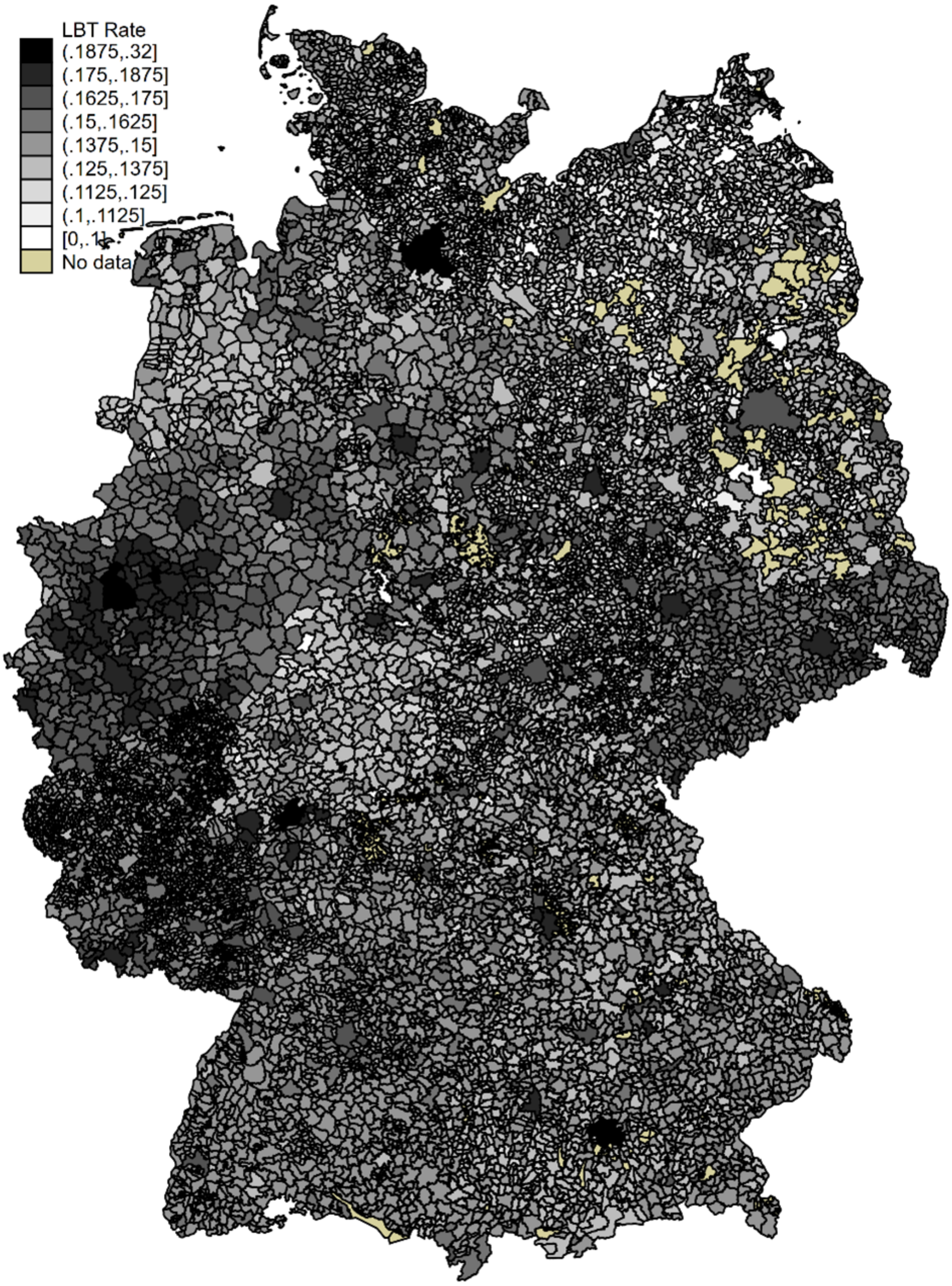


Figure 2: Spatial Distributions of LBT Rates (2006)

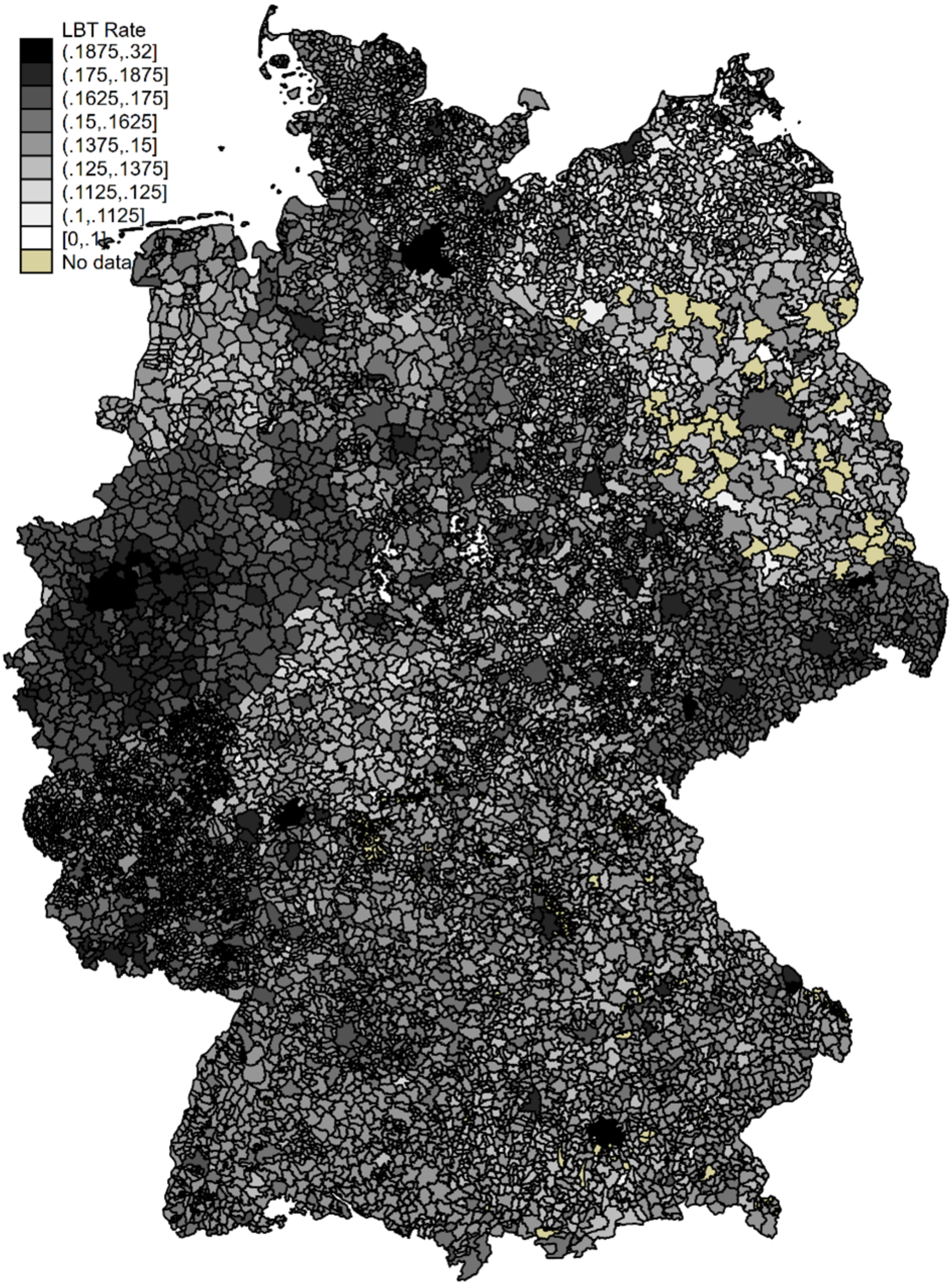


Figure 3: Geographic Location of Tax Havens within Germany (2002)

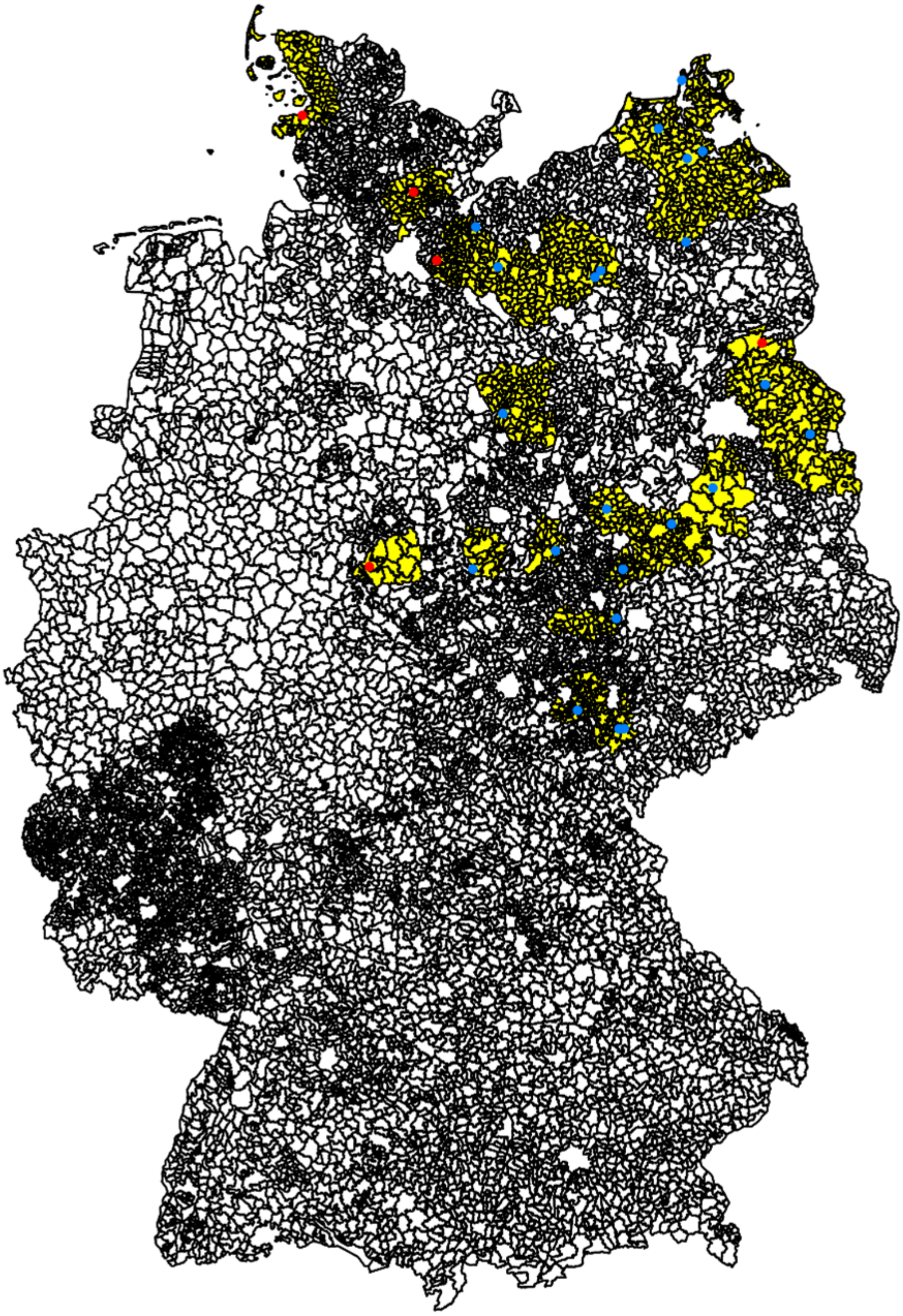


Figure 4: Employees per Inhabitant (2001-2006)

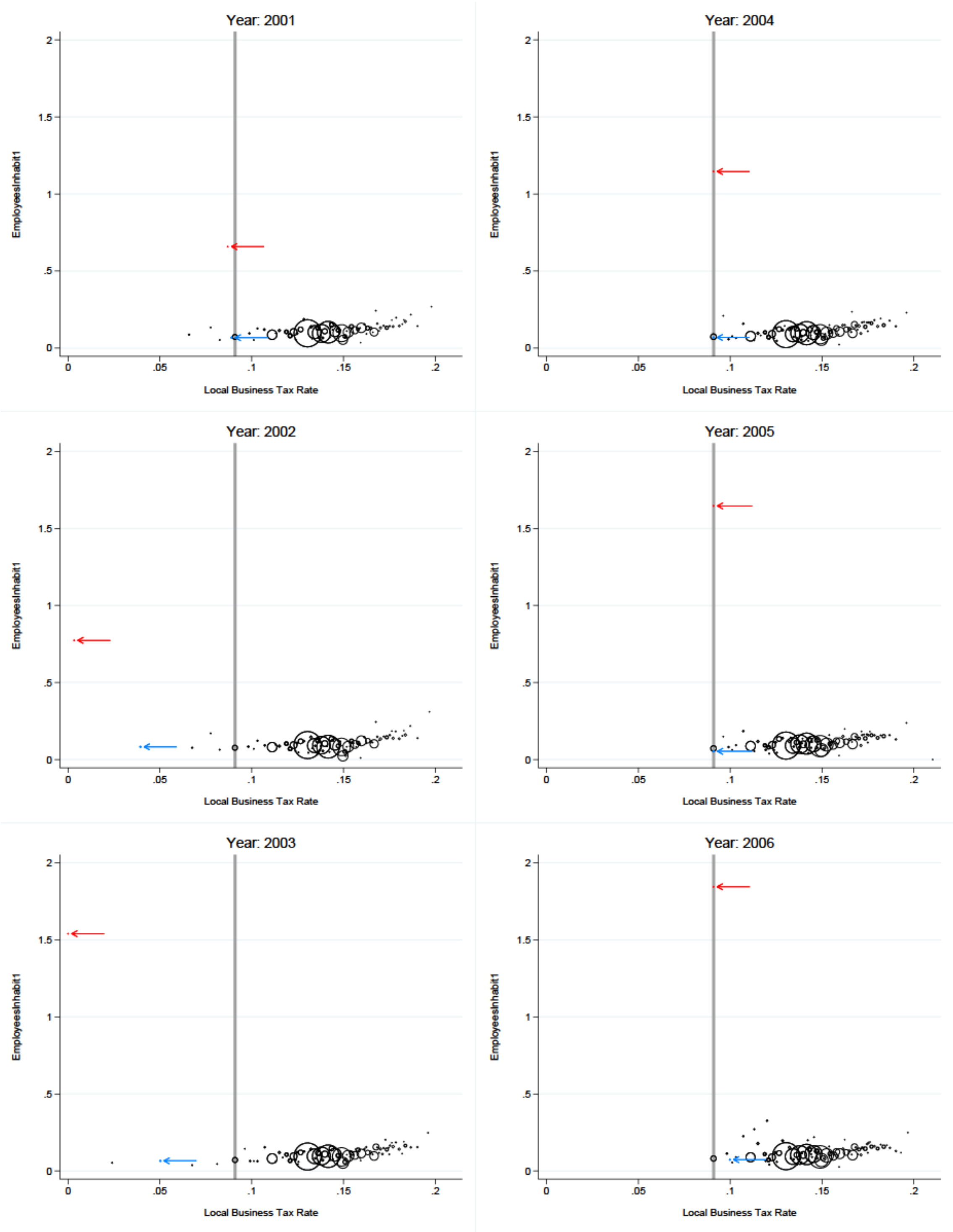


Figure 5: Employees per Firm (2001-2006)

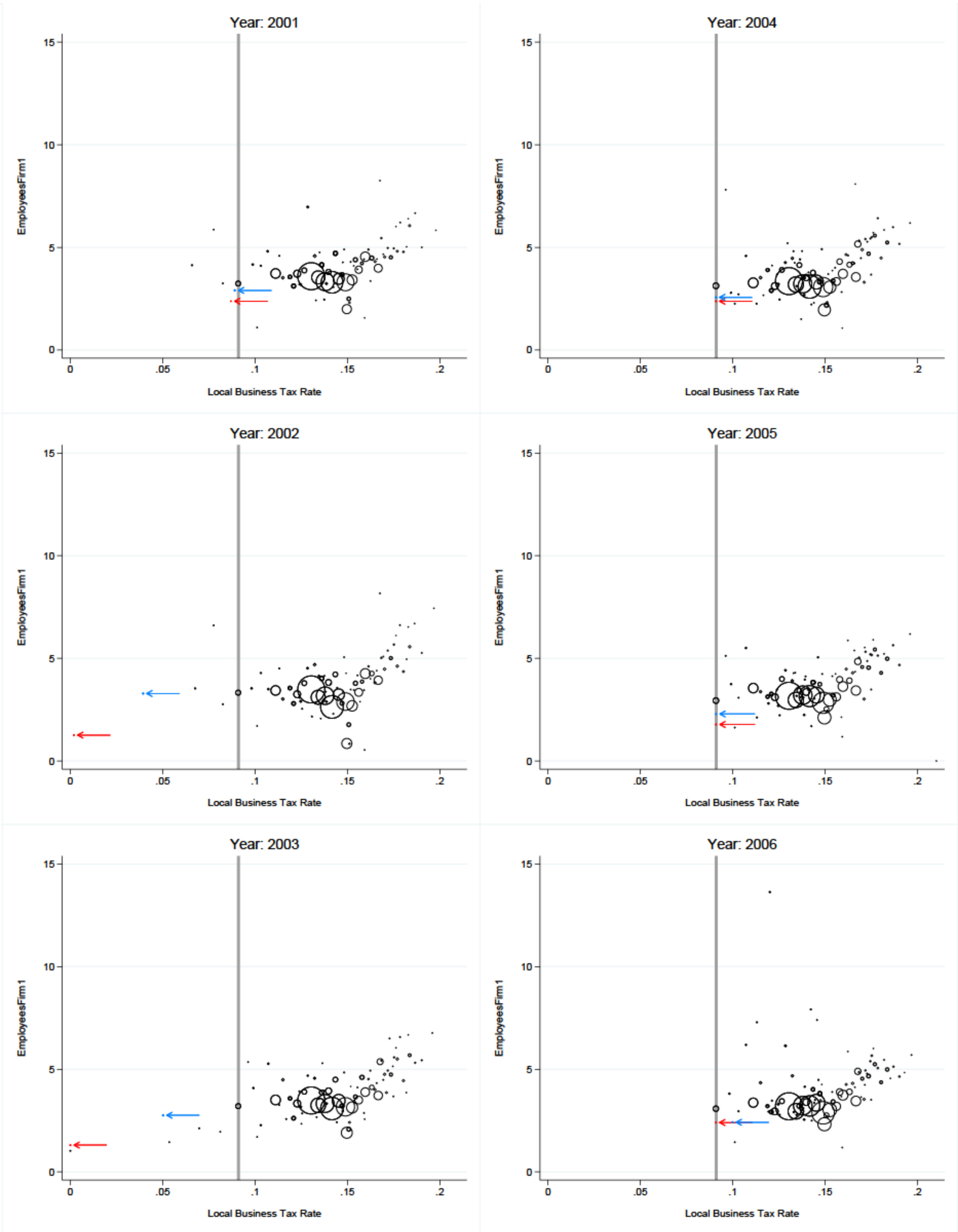


Figure 6: Estimated LBT Base per Inhabitant (2001-2006)

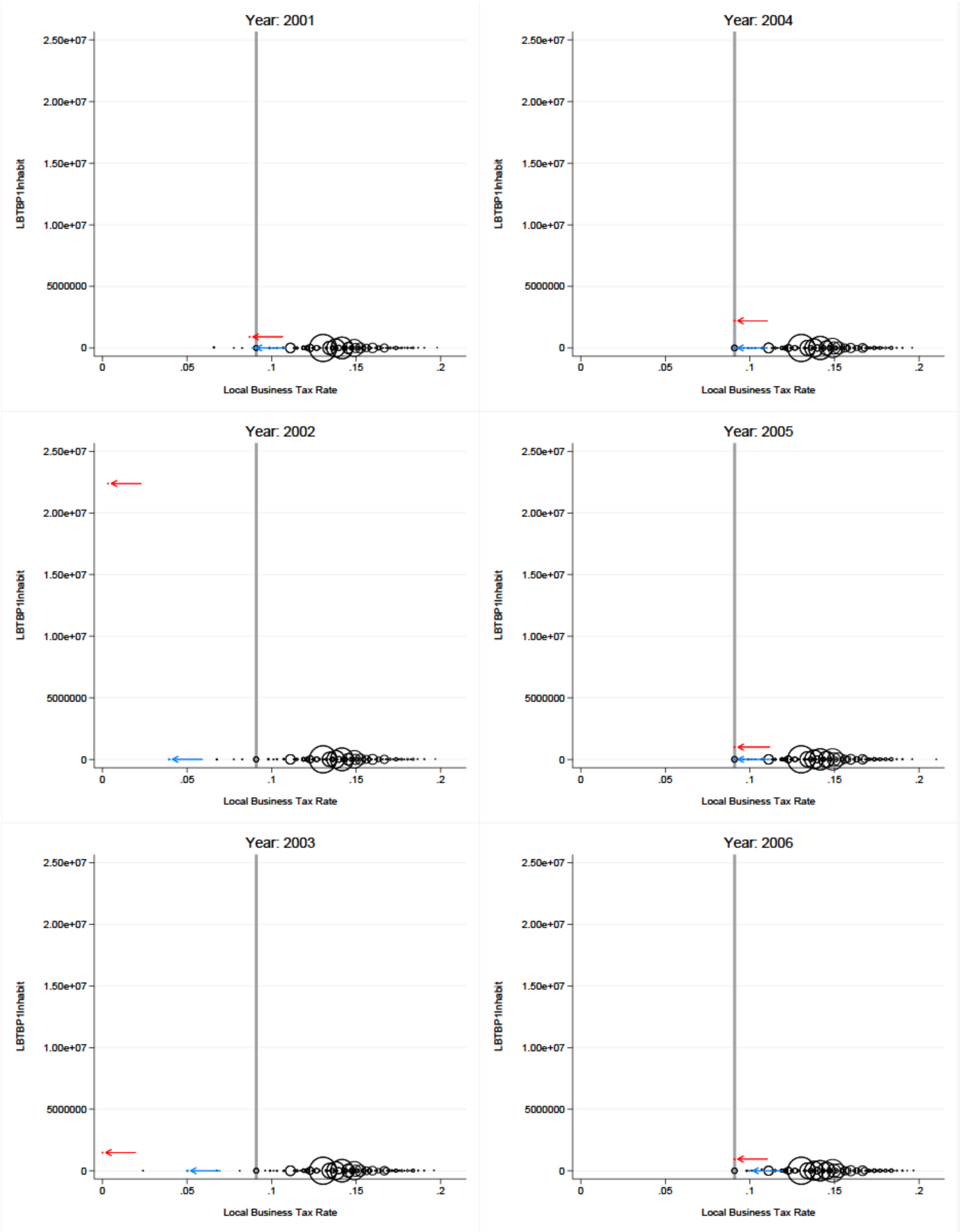


Figure 7: Net Sales per Inhabitant (2001-2006)

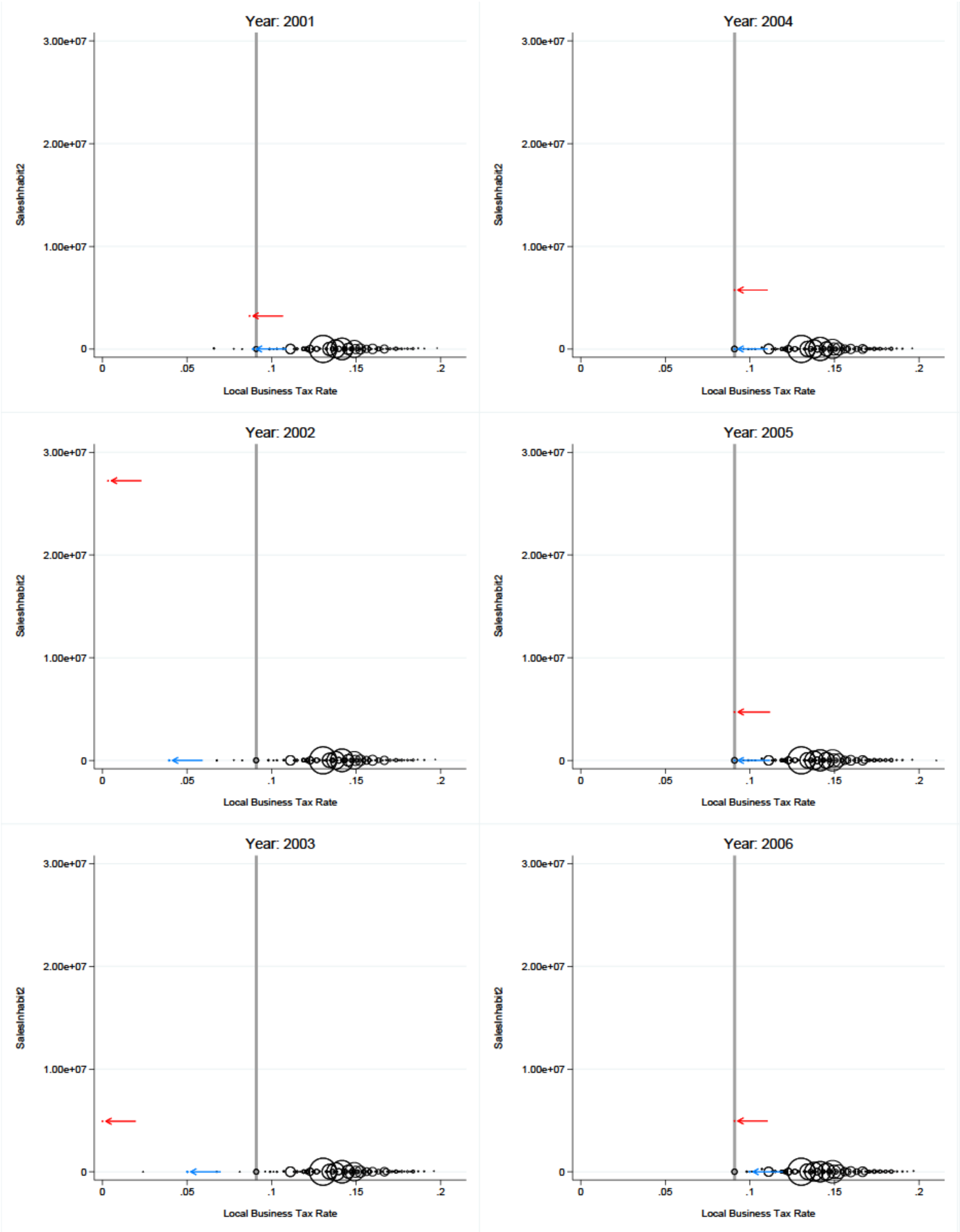


Figure 8: Actual LBT Amount per Inhabitant (2001-2006)

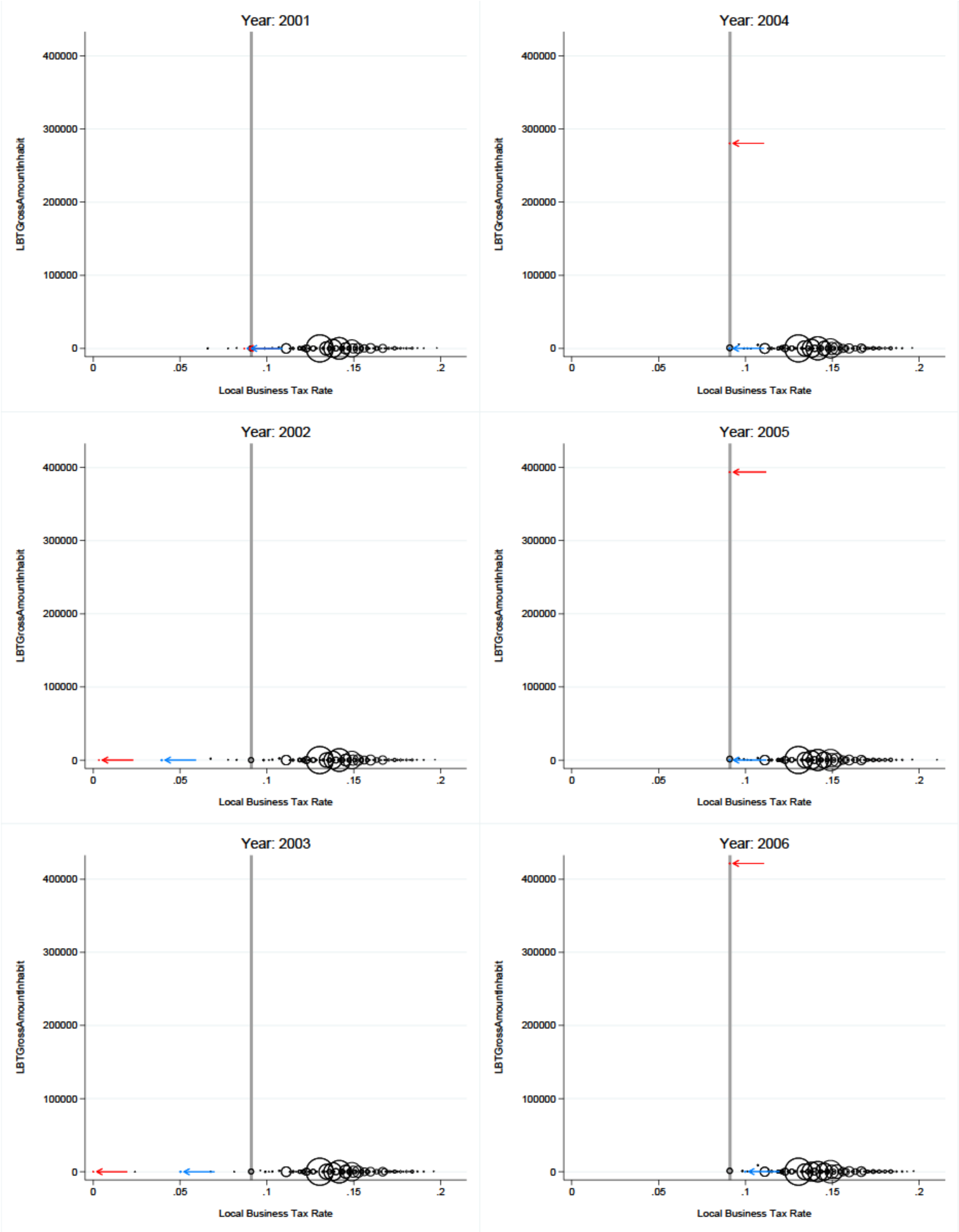


Figure 9: Industry Breakdown for Tax Havens and Similar Non-Havens

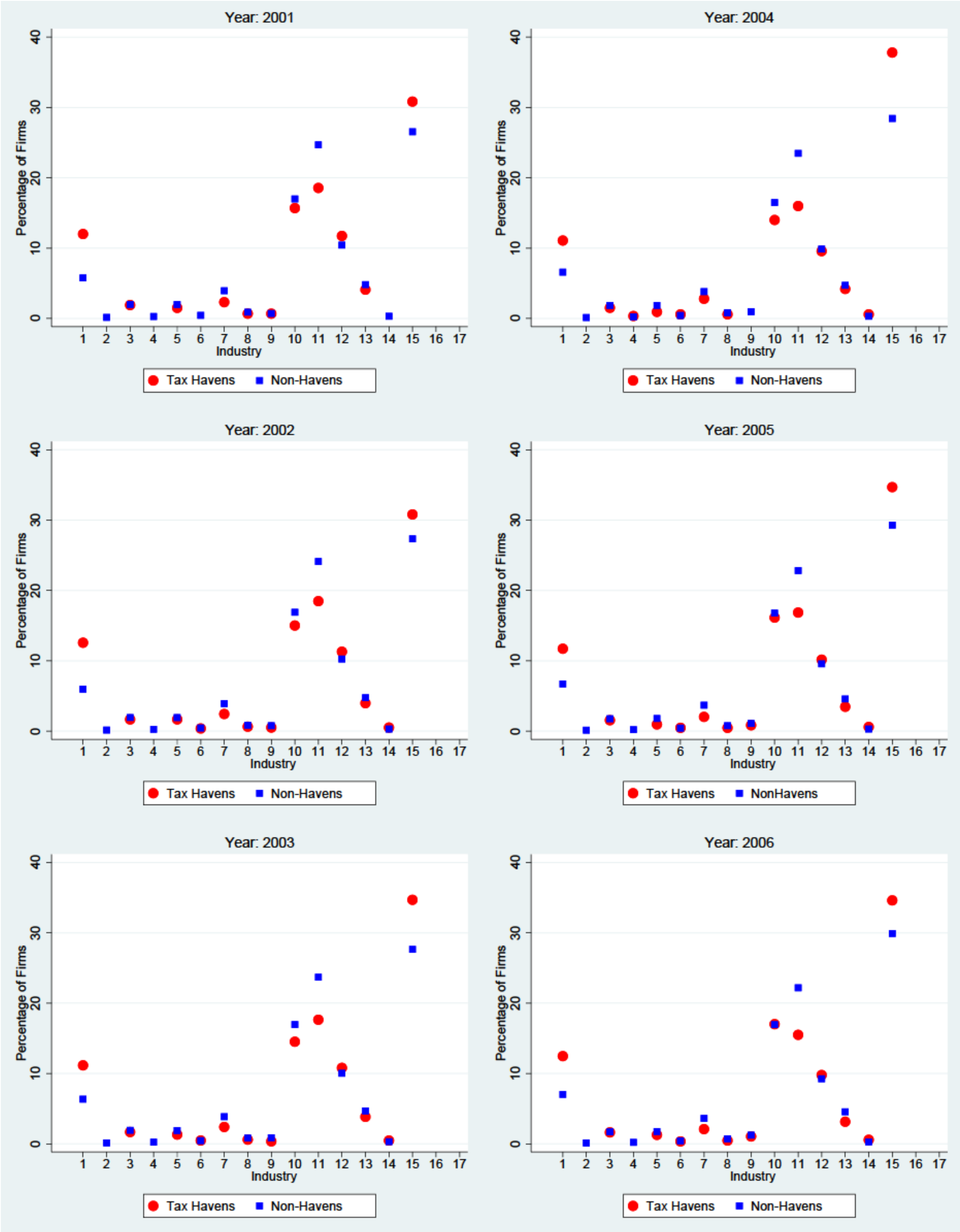


Table 1: Descriptive Statistics for Tax-Havens and Non-Havens (2001-2003)

				Quartiles		
	N	**Mean**	**Std-Dev**	**25%**	**50%**	**75%**
Panel A: Tax Havens						
Population	74	1,083	1,588	286	613	1,173
Δ Population	49	-3.78	36.85	-16.00	-4.00	3.00
Population/km²	74	47.62	52.97	21.62	31.20	51.57
Surface Area (km²)	74	36.29	68.82	6.57	15.03	29.69
Farming Area (%)	62	56.75	28.30	37.19	68.04	76.06
Business Area (%)	41	0.35	0.75	0.00	0.08	0.16
Land Value (€/m²)	20	22.12	10.73	17.09	21.75	26.14
Unemployment Rate (%)	43	8.72	3.09	7.12	8.56	10.65
GDP/Capita, district (€)	32	16,760	3,683	14,226	14,811	19,214
Panel B: Similar Non-Havens						
Population	31,655	2,103	2,215	529	1,194	2,927
Δ Population	21,018	4.71	92.84	-12.00	-1.00	13.00
Population/km²	31,655	124.85	145.40	47.42	81.27	145.51
Surface Area (km²)	31,655	21.40	22.10	7.40	14.67	27.27
Farming Area (%)	30,709	59.25	20.71	44.94	60.88	75.90
Business Area (%)	25,912	0.58	1.15	0.07	0.26	0.63
Land Value (€/m²)	19,549	68.85	65.12	28.45	46.29	84.05
Unemployment Rate (%)	20,951	5.16	3.53	2.59	3.67	7.38
GDP/Capita, district (€)	25,062	19,663	4,802	16,626	19,816	22,162
Panel C: All Non-Havens						
Population	36,164	6,754	44,105	602	1,517	4,558
Δ Population	24,027	13.34	300.37	-13,00	0.00	16.00
Population/km²	36,164	180.41	266.69	52.44	93.38	189.55
Surface Area (km²)	36,164	27.72	33.49	8.21	16.90	34.24
Farming Area (%)	35,204	57.90	20.58	43.61	59.14	74.23
Business Area (%)	30,239	0.80	1.44	0.10	0.34	0.88
Land Value (€/m²)	23,636	76.35	73.14	30.46	51.08	95.84
Unemployment Rate (%)	23,982	5.06	3.41	2.64	3.67	6.92
GDP/Capita, district (€)	29,181	20,259	5,681	16,909	20,146	22,450

Table 2: LBT Rate Changes by Year

	2002	2003	2004	2004*	2005	2006	Difference between 2004 and non-2004 average	Difference between 2004 and non-2004 average*
Rate increases, % of municipalities	7.5	8.86	8.08	7.91	10.33	7.34	-0.43	-0.59
Mean change conditional on an increase, percentage points	0.78	0.8	0.91	0.81	0.75	0.75	0.14	0.03
Median change conditional on an increase, percentage points	0.67	0.74	0.73	0.73	0.73	0.38	0.1	0.1
Rate decreases, % of municipalities	0.67	0.65	0.62	0.62	0.69	0.86	-0.1	-0.1
Mean change conditional on a decrease, percentage points	-1.71	-1.45	-1.36	-1.36	-1.16	-1.02	-0.02	-0.02
Median change conditional on a decrease, percentage points	-0.73	-0.38	-0.75	-0.75	-0.73	-0.72	-0.11	-0.11
Mean change conditional on a change, percentage points	0.579	0.65	0.744	0.648	0.626	0.563	0.139	0.043
Mean change, percentage points	0.047	0.062	0.065	0.055	0.069	0.046	0.009	-0.0008
Municipalities	12083	12057	12109	12109	12147	12160		

Note: * indicates considering 2003 and 2004 jointly, i.e., treating 2003 as an additional “treated” year.

Table 3: OLS Regressions of LBT Rates on Weighted Average of Others' LBT Rates

	Closest 3 by midpoint distance	**Closest 5 by midpoint distance**	**All in same district**	**Inverse distance**	**Inverse distance squared**	**Population/ distance**	**Population/ distance squared**	**Population/ distance squared to 10 biggest citites**
				2001-2003				
Neighbors' average rate	0.404***	0.500***	0.735***	1.103***	0.322*	-0.04	0.266***	0.007
Standard error	0.006	0.007	0.01	0.047	0.008	0.034	0.012	0.016
Observations	34699	34765	34438	34585	34585	34780	34780	34780
R-squared adj	0.5871	0.5956	0.5787	0.5396	0.5527	0.5322	0.5387	0.5322
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-by-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
				2004-2006				
Neighbors' average rate	0.456***	0.546***	0.820***	1.152***	0.328***	0.026	0.200***	0.009
Standard error	0.006	0.007	0.011	0.049	0.008	0.033	0.01	0.016
Observations	35916	35976	35655	35782	35782	35991	35991	35991
R-squared adj	0.6082	0.6141	0.6017	0.556	0.5684	0.5487	0.5534	0.5487
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-by-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*** $p<0.01$, ** $p<0.005$, * $p<0.1$.

Table 4: 2SLS Regressions Instrumenting for Neighbors' LBT Rates with Neighbors' Demographics

	Closest 3 by midpoint distance	**Closest 5 by midpoint distance**	**All in same district**	**Inverse distance**	**Inverse distance squared**	**Population/ distance**	**Population/ distance squared**	**Population/ distance squared to 10 biggest citites**
				2001-2003				
Neighbors' average rate	0.202***	0.240***	0.564***	0.808***	0.111***	0.374***	0.045***	0.079***
Standard error	0.029	0.029	0.036	0.038	0.012	0.028	0.018	0.021
Observations	21386	22310	22308	22513	22513	22649	22649	22649
R-squared	0.5953	0.5979	0.5901	0.5524	0.5572	0.541	0.5475	0.5451
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-by-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
				2004-2006				
Neighbors' average rate	0.337***	0.389***	0.638***	0.784***	0.094***	0.439***	0.035**	0.028*
Standard error	0.03	0.03	0.042	0.033	0.012	0.026	0.019	0.021
Observations	23356	24336	24180	24356	24356	24515	24515	24515
R-squared	0.6081	0.6099	0.5994	0.5542	0.5578	0.5433	0.5494	0.5476
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-by-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*** p<0.01, ** p<0.005, * p<0.1.

Table 5: 2SLS Regressions Instrumenting for Neighbors' LBT Rates with Increases Required to Comply with Minimum Rate (in Levels)

	Closest 3 by midpoint distance	**Closest 5 by midpoint distance**	**All in same district**	**Inverse distance**	**Inverse distance squared**	**Population/ distance**	**Population/ distance squared**
			2001-2003				
Neighbors' average rate	0.043***	0.041**	0.249***	0.817***	0.102***	1.022***	0.853***
Standard error	0.016	0.02	0.026	0.152	0.031	0.033	0.026
Observations	34625	34696	34369	34529	34529	34724	34724
R-squared	0.5559	0.5547	0.566	0.5519	0.5559	0.531	0.5188
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-by-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
			2004-2006				
Neighbors' average rate	0.105***	0.115***	0.368***	1.365***	0.276***	1.088***	0.957***
Standard error	0.028	0.033	0.038	0.216	0.065	0.031	0.031
Observations	35546	35822	35501	35628	35628	35837	35837
R-squared	0.5771	0.5772	0.5867	0.5594	0.5717	0.5392	0.4916
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-by-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Sample excludes municipalities forced to raise rates by the minimum.

Minimum instrument includes no variation in the tax rates of the ten largest citites.

*** p<0.01, ** p<0.005, * p<0.1.

Table 6: 2SLS Regressions of 2002-2004 Changes in own LBT Rate, Instrumenting for Neighbors' LBT Rates with Increases Required to Comply with Minimum Rate

	Closest 3 by midpoint distance	**Closest 5 by midpoint distance**	**All in same district**	**Inverse distance**	**Inverse distance squared**	**Population/ distance**	**Population/ distance squared**
Change in neighbors' average tax rate	-0.007	-0.009	-0.014	0.374	0.039	73.244	0.456
Standard error	0.018	0.022	0.036	0.473	0.047	96.457	0.47
Observations	34625	34696	34369	34529	34529	34724	34724
R-squared	0.018	0.022	0.036	0.473	0.047	96.457	0.47
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-by-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Sample excludes municipalities forced to raise rates by the minimum.

Minimum instrument includes no variation in the tax rates of the ten largest citites.

*** $p<0.01$, ** $p<0.005$, * $p<0.1$.

Acknowledgments

We thank Shafik Hebous, Petro Lisowsky, Dominika Langenmayr and Reinald Koch, and the participants at the National Tax Association Annual Conference on Taxation 2017, Annual Symposium of the Centre of Business Taxation 2018, Conference on Accounting, Auditing and Taxation Siegen 2018, Annual MaTax Conference 2018, the Augolstadt Seminar 2021, and the 23rd Annual Conference on Finance and Accounting for their 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.

Data Source

RDC of the Federal Statistical Office and the Statistical Offices of the Federal States, VAT Panel, 2001-2006 (DOI: 10.21242/73311.2017.00.01.2.1.0).

Norderfriedrichskoog! Tax Havens, Tax Competition and the Introduction of a Minimum Tax Rate

Online Appendix A.

Combining the VAT data with data on municipalities' public finances, we construct a proxy for the LBT base on a firm-year basis. As the LBT has a flat rate and only a small exemption amount for small firms, dividing a municipality's annual LBT revenues by its LBT rate creates a highly accurate measure of the LBT base. However, in tax havens where the LBT rate and revenues are zero, this method is not available. In these cases, we proxy for the LBT base in a municipality with firms' total sales net of non-labor input costs. This proxy differs from the true LBT base mainly because of 1) exemptions for small firms, 2) partial non-deductibility of interest expense in the computation of LBT base, and 3) labor costs and depreciation being subtracted from the actual LBT base but not from the VAT base. However, the proxy is still highly correlated with the actual LBT base and, most importantly, can be computed even for zero rate tax havens. The proxy can naturally be computed at the firm-year level in our data.

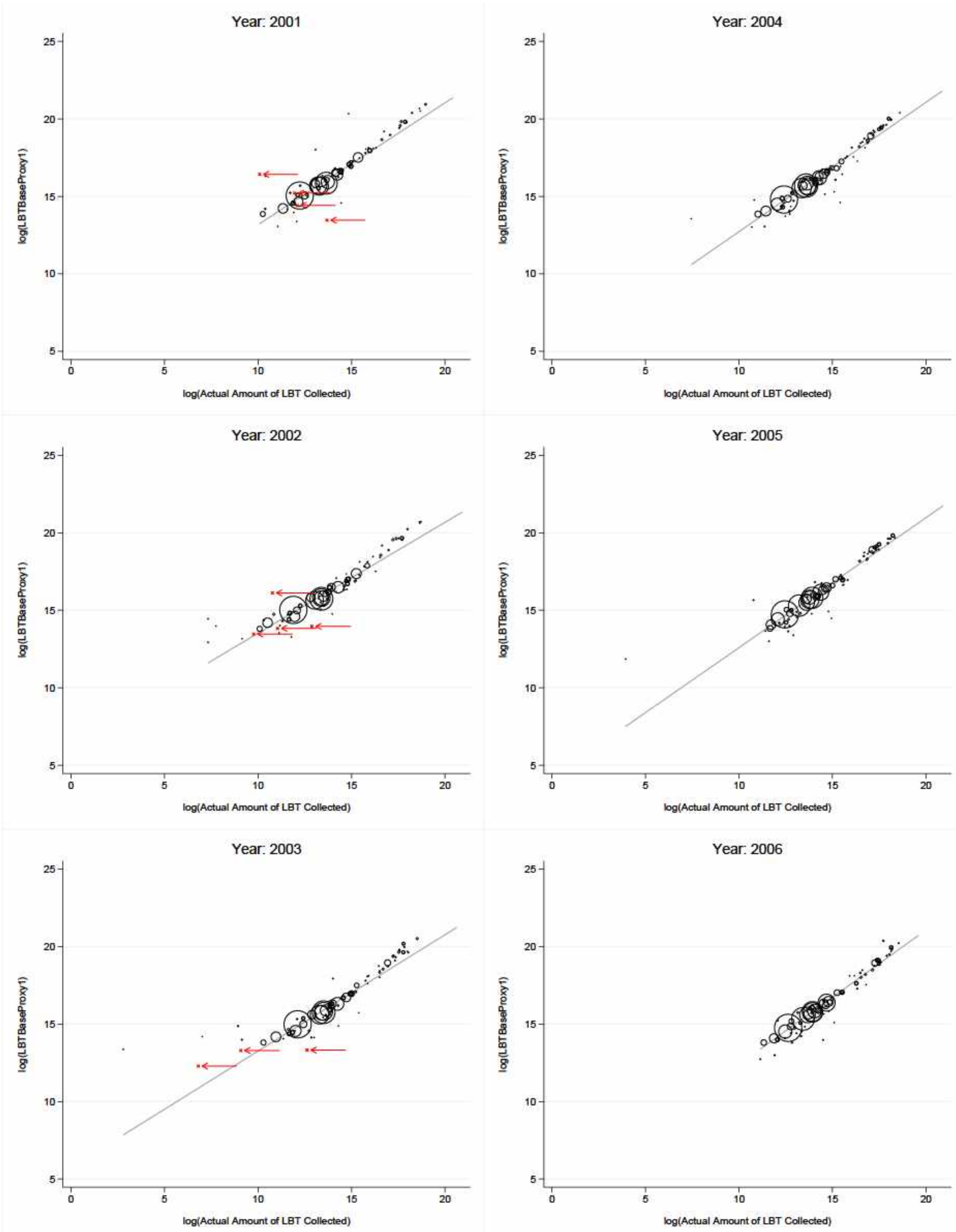
To obtain the proxy, we would theoretically need to subtract the LBT-deductible input costs from sales. Sales are directly available on VAT returns, whereas input costs are not. Consequently, we need to calculate input costs from the input VAT paid, which requires that we use a (value added) tax rate applicable for each firm-year. The VAT rate to be used here is not easily defined, as Germany had two tax rates, 7% and 16%, in the period of investigation, and both (or a mixture) could be applied to input costs of firms in our sample.

We then test the viability of our proxy by using the total amount of our proxy over all firms per municipality-year and test its correlation against the LBT base computed from LBT revenue per municipality. Such a test is obviously limited to those municipalities in which the actual LBT base could be estimated from the actual LBT revenue reported, which is the case for approximately 65% of the municipalities over all years.

Our first proxy for the tax base treats 16% as the average VAT rate to apply when calculating deductible input costs. Hence, it is the most conservative measure of non-labor input costs. By this first proxy for the tax base, in most years, the lowest-LBT rate havens had more than €1000,000 of tax base per inhabitant, and in 2002 over €20,000,000 per inhabitant. The proxy is correlated with the estimate of the actual LBT base with a correlation coefficient of 0.816. Figure A.1 below illustrates that the proxy performs well across computed LBT base levels and LBT rates.

Specifically, we order municipalities by their LBT rate per year and form bins of equal tax rates. When any such bin does not contain at least three municipality observations, which is the minimum confidentiality requirement for bins to be reported, we join this bin with the next highest LBT-rate bin until the joint bin contains at least three municipality observations. The size of the bubble indicates the number of municipalities that are placed in each LBT-rate bin; the red crosses with indicator arrows mark bins that have an average LBT rate lower than 9.1% in the respective year. Values are shown in natural logarithm adding one in all cases to avoid logging zeros.

Figure A.1: Correlation of Actual and Estimated LBT Base (2001-2006)



We report that the fitted line, in general, seems to fit well with the underlying data, and hence, that our proxy is strongly correlated with the LBT base at the municipality level. This again indicates that it is a good proxy at firm level. Specifically, regarding the municipalities with tax rates lower than the minimum rate, we find that these are also near the fitted regression line. We also note that we do not include tax havens with exactly zero tax rates from 2001 to 2003, as these would not provide us with an estimate of firm-level LBT base from LBT collected at the municipality level, that is, we expectedly observe no values of exactly zero on the x-axis.

A second proxy uses a VAT rate calculated as the average VAT rate paid over the entire universe of German firms, treating the data as cross-sectional. This implies that the VAT rate paid on any input cost for each firm corresponds to the average VAT rate in Germany. The proxy is correlated with the estimate of the actual LBT base with a correlation coefficient of 0.620. A third proxy computes the input VAT average tax rate similarly, but on a per-year basis. The proxy is correlated with the estimate of the actual LBT base with a correlation coefficient of 0.609.

For the main analyses, we use the first proxy for our estimation of the LBT base on the level of the firm and municipality.

Essay 3

The Effects of Scandals on Organizational Affiliation and Competition: Evidence from Church Scandals in Germany

The Effects of Scandals on Organizational Affiliation and Competition: Evidence from Church Scandals in Germany

Drahomir Klimsa*, Erik E. Lehmann†, Robert Ullmann‡, Laurenz Weiße§

Version Date: May 30, 2022.

Abstract

This paper investigates the effects of scandals on organizations and their stakeholders. We introduce a novel framework that links the conceptual origin of a scandal, i.e., individual-caused vs. institution-caused, with its impact on affiliation with the scandal-stricken organization and with the scandal-stricken organization's competitors. In our analysis, we exploit the changes in diocese-level and regional church-level measures of affiliation with the Catholic Church and Protestant Church in Germany that followed numerous scandals involving the two major German religious organizations between 2002 and 2016. We find that both individual-caused and institution-caused scandals are associated with a decline in affiliation with the scandal-stricken organization. However, individual-caused scandals have a significantly larger effect on affiliation with the scandal-stricken organization than institution-caused scandals. We also find evidence of positive interorganizational spillover effects on unassociated competitors of the scandal-stricken organization, but only for institution-caused scandals. Our results contribute to the emerging field of studies on the effects of scandals on organizations and their stakeholders. Moreover, due to the economic character of religious organizations, i.e., because they compete in a religious market to provide pastoral care, our results can be generalized beyond our empirical setting to secular organizations and their stakeholders.

Keywords: church tax, organizations, religion, reputation, scandals

JEL Classification: J11, L14, L30, M14, Z12

*University of Augsburg.

†University of Augsburg.

‡University of Augsburg.

§University of Augsburg.

1. Introduction

Organizations are not free of misconduct by the organization's employees or the organization's management. When revealed to the public, organizational misconduct often transforms into a scandal with potentially wide-ranging consequences for the scandal-stricken organization and the scandal-stricken organization's stakeholders. Although a considerable number of organizational scandals can be cited from recent economic history, e.g., scandals involving Wirecard, Volkswagen, Parmalat, Worldcom, or Enron, the effects of a scandal on the scandal-stricken organization and scandal-stricken organization's stakeholders are still not well understood. While prior research provides insights into the effects of scandals on the scandal-stricken organization's shareholders, mostly considering stock market reactions, less is known about how scandals affect other key organizational stakeholders, particularly individuals affiliated with the scandal-stricken organization, such as members, employees, or customers, and the scandal-stricken organization's competitors. The limited evidence likely stems mainly from data availability issues, as it is challenging to collect information about a sufficiently large sample of scandals for an empirical analysis and to obtain information about individuals affiliated with the scandal-stricken organization and the scandal-stricken organization's competitors due to its private character. To our knowledge, Piazza and Jourdan (2018) are among the first to provide large-scale evidence regarding the effects of scandals on organizational affiliation and competition. Analyzing the effects of scandals on church membership in the U.S. Catholic Church after the publication of scandals related to sexual abuse by Catholic clergy, their results indicate that scandals are positively associated with a decline in membership. Furthermore, they find positive spillover effects on non-Catholic religious organizations that benefit from the scandals via an increase in membership.

In this paper, we use genuine diocese-level and regional church-level data on affiliation made available by the two major religious organizations in Germany, i.e., the Catholic Church and Protestant Church, to provide further insights into the effects of scandals on organizational affiliation and competition. Specifically, we analyze the effects of 137 hand-collected locally reported scandals that occurred in Germany's 27 Catholic dioceses and 19 Protestant regional churches between 2002 and 2016 to examine the effects of scandals on affiliation with the scandal-stricken organization and the scandal-stricken organization's competitors. Our analysis goes beyond prior research in terms of at least two major aspects. First, we use a broad range of measures of affiliation beyond membership, specifically, Sunday service attendance rate, exit rate, entry rate, and baptism rate, which allows us to study the short-term and long-term effects of scandals and different degrees of (dis)affiliation. Second, as the scandals in our setting are not limited to sexual abuse scandals but also include various other scandal-categories, such as scandals related to financial misconduct, church as an employer, and church as a public service provider, this allows us to consider the different effects of scandals with different conceptual origins. Specifically, we differentiate between scandals related to misconduct directly connected to a poor governance system, i.e., institution-caused scandals, and scandals related to misconduct with no indications that the organization's governance system enabled, facilitated or even encouraged the scandalous behavior, i.e., individual-caused scandals.

We find that both individual-caused and institution-caused scandals are associated with a decline in affiliation with the scandal-stricken organization. However, individual-caused scandals have a significantly larger effect on affiliation with the scandal-stricken organization than institution-caused scandals. We also find evidence for positive interorganizational spillover effects on unassociated competitors of the scandal-stricken organization but only for institution-caused scandals.

Our study contributes mainly to two strands of the literature. First, although it is not our primary objective, we contribute to studies on religious disaffiliation by investigating the effects of a large number of scandals on affiliation with the German Catholic Church and German Protestant Church. Second, we contribute to the emerging field of studies on the effects of scandals on organizations and their stakeholders. Although there are naturally many differences between religious organizations and secular organizations, we argue that due to the economic character of religious organizations (religious organizations compete on the religious market in offering pastoral care), our results can be generalized beyond our empirical setting, particularly to stakeholders of secular organizations, i.e., members of political parties, trade unions, or NGOs, as well as employees and customers of private firms.[1] This view is consistent with the literature on the economics of religion, which considers the Catholic Church to be the oldest and most enduring organization in the Western world and a multidivisional firm (Rost, 2017).

1 As membership in religious organizations is usually passed down from parents to their children (Frick and Simmons, 2017), and members of religious organizations consequently have particularly high membership benefits compared to most secular organizations due to social ties to other members of the organization, we argue that the effects of scandals observed in our setting represent the lower bound estimate of the effects of scandals on organizations and their stakeholders.

The remainder of this paper is organized as follows: Section 2 reviews the previous literature. Section 3 provides information on the institutional setting, and Section 4 outlines our theory and develops our hypotheses. Section 5 describes the empirical strategy and Section 6 introduces our data. Section 7 presents the empirical evidence, and Section 8 concludes the paper.

2. Literature Review

2.1. Definition and Typology of Scandals

Although there is currently no generally accepted typology of how to classify scandals, a common understanding in the related research is that scandals are the result of real or alleged misconduct, i.e., a normative violation that is invoked by the transgression of the line separating legal and legitimate from illegal and illegitimate (Tarrow, 1994; Adut, 2008; Thompson, 2013). Regardless of whether scandals are based on true or false allegations, a defining element of a scandal is a social reaction of outrage, anger, or surprise that occurs or spreads at the moment an allegation is publicized (Thompson, 2013). While misconduct always depicts the start of a scandal, not every transgression of societal rules necessarily turns into a scandal, as some amount of media attention is needed to transform misconduct into a scandal. Specifically, the media act as a funnel and a catalyst that "*discerns between a transgression that progresses into a scandal and the many transgressions that remain buried in the mass of daily news*" (Clemente and Gabbioneta, 2017, pg. 288).

Against this background, one strand of literature in the organizational studies literature divides scandals based on the type of misconduct underlying the scandal, such as fraud, product (safety) issues, employee mistreatment, or environmental violations (e.g., Hersel et al., 2019). Another strand of literature divides scandals by their controllability, severity, undesirability, and intentionality (e.g., Coombs, 1995; Bundy and Pfarrer, 2015; Wooten and James, 2008; Lange and Washburn, 2012). A third strand of literature points out the importance of the (conceptual) origin of a scandal (Lehmann, 2019; Linstead et al., 2014), i.e., the role of the organization in the occurrence of the misconduct underlying the scandal. Specifically, even though misconduct is usually viewed as the result of an individual decision-making process, i.e., an individual within an organization engages in misconduct when the (expected) benefits of the action outweigh the (expected) costs of the action (Draca and Machin, 2015; Lehmann, 2012), the individual decision-making process is influenced not only by individual factors but also by the specific organizational environment. Thus, according to this literature, scandals must be examined individually and institutionally. In the fourth and final strand of literature, scandals are differentiated by the affected party of the scandal, i.e., whether the effects are limited to the scandal-stricken organization itself or whether there are spillover effects on other organizations (Hersel et al., 2019), in particular, the organization's competitors.

2.2. Effects of Scandals on Scandal-Stricken Organizations

Most prior literature has focused on the effects of scandals on the scandal-stricken organization's shareholders by analyzing stock-market reactions to scandals. From a stock-market perspective, the change in a firm's stock price after the occurrence of a scandal reflects the reevaluation in investor expectations about future firm performance (Karpoff et al., 1999). For instance, using a sample of scandals related to managerial fraud (i.e., insider trading, corruption, cartel agreements, embezzlement, and accounting fraud) of listed firms in Germany between 1998 and 2004, Ewelt-Knauer et al. (2015) show a significant (short-term) decline in stock prices after the occurrence of misconduct becomes public and transforms into a scandal. Furthermore, Song and Han (2017) analyze stock market reactions to scandals related to different types of misconduct. They show that scandals related to financial misconduct have a significantly stronger negative impact on stock prices than scandals related to other types of organizational misconduct. On the other hand, there is also some evidence for misconduct that transforms into a scandal having positive effects on shareholders. Specifically, examining long-term stock market returns, i.e., before and after the public disclosure of misconduct, particularly managerial fraud, Tibbs et al. (2011) provide evidence of a net benefit of misconduct for shareholders if the party damaged by the misconduct is unrelated to the scandal-stricken organization. More specifically, they find a stock market outperformance in the period before disclosure that is only partially reversed by the decline in stock prices in the period following the disclosure of misconduct.

In addition to the effects of scandals on shareholders, previous research has shown negative effects of scandals on customers of scandal-stricken organizations. Specifically, prior literature suggests a general decline in purchase intention and product evaluation and an increase in negative customer opinions (Byun et al., 2020; Monga and John, 2008; Wang and Alexander, 2018) following scandals related to product recalls. These effects are especially pronounced for scandal-stricken organizations with well-established brand awareness (Korkofingas and Ang, 2011; Zavyalova et al., 2016). Furthermore, Rhee and Haunschild (2006) find that the negative effects of product recalls on customers are stronger for highly reputed firms than for less reputable firms.

Prior research has also shown negative effects of scandals on donations. Examining scandals related to organ donations in Germany in 2012 and 2014, Röck et al. (2017) find a significant decline in donations in the year after the scandals. Furthermore, negative effects on financial donations to the U.S. Catholic Church were shown by Bottan and Perez-Truglia (2015) following scandals related to sexual abuse by Catholic clergy.

Prior literature has also shown negative effects of scandals on employees. Specifically, scandals related to employee discrimination or sexual harassment at the workplace increase employee turnover rates in scandal-stricken organizations (Goldman et al., 2006; Madera et al., 2012; Del Triana et al., 2019). Furthermore, using employees' ratings of S&P 500 firms, Zhou and Makridis (2019) find a significant decline in employees' perceptions of their employers after scandals related to accounting fraud. Similar effects were found by Gadgil and Sockin (2020), who identify an immediate and long-lasting negative impact on employee perception following scandals. Specifically, they show that the likelihood of employees recommending their employer on social networks decreases after a scandal.

Another strand of literature closely related to our work has shown negative effects of scandals on members of (religious) organizations. For instance, Hungerman (2013) shows a decline in membership in the U.S. Catholic Church following scandals related to sexual abuse by Catholic clergy during the early 2000s. Furthermore, Bottan and Perez-Truglia (2015) show that the effects of such scandals are mostly concentrated in the areas in which the scandals occur. Moreover, Piazza and Jourdan (2018) find a positive association between the publicity of such scandals and the decline in membership.

2.3. Effects of Scandals Scandal-Stricken organizations' Competitors

In addition to effects on scandal-stricken organizations, their shareholders and other stakeholders affiliated with the scandal-stricken organization (customers, employees, and members), prior research has shown that scandals can also have considerable effects on organizations not involved in the scandal, particularly competitors of the scandal-stricken organization. Negative spillover effects can occur if stakeholders of a competitor of the scandal-stricken organization or the public perceive that the competitor may be involved in misconduct due to its (perceived) similarity to the scandal-stricken organization (contamination). Specifically, negative spillover effects occur if the specific competitor is perceived to be associated "*personally, institutionally, or even categorically*" (Adut, 2008, pg. 24) with the scandal-stricken organization.

For instance, Bouzzine and Lueg (2020) find immediate negative spillover effects on the stock market returns of the main competitors of VW after VW's "Dieselgate" scandal. Furthermore, a study by Paruchuri and Misangyi (2015) on scandals related to financial misconduct confirms such negative effects on the stock market returns of competitors of the scandal-stricken organization. Moreover, Jonsson et al. (2009) show negative effects of scandals related to asset exploitation and unjustified incentive schemes on other mutual fund providers analyzing scandals involving the Swedish mutual fund provider Skandia AB.

Although prior research has predominantly focused on the negative spillover effects of scandals, there is also growing evidence on the positive spillover effects of scandals for competitors of scandal-stricken organizations, as competitors who offer similar goods and services may attract some of the scandal-stricken organizations' customers (substitution), and, thus improve their competitive position (Piazza and Jourdan, 2018). Specifically, positive spillover effects on firm sales were shown after scandals related to toy recalls by Ni et al. (2016). Furthermore, utilizing the scandal related to Tiger Woods' extramarital affairs in 2009, Knittel and Stango (2014) show negative stock market returns for Tiger Wood's major sponsors and (short-term) positive effects on the stock market returns of competitors not involved in celebrity endorsement. Finally, Hungerman (2013) and Bottan and Perez-Truglia (2015) find positive effects on membership in competing religious organizations following scandals related to sexual abuse in the U.S. Catholic Church. Piazza and Jourdan (2018) show that these positive spillover effects are particularly concentrated on organizations with similar offerings but with a stricter organizational code of conduct.

3. Institutional Setting

In most countries, historically grown constitutional norms and treaties between the respective country and its religious organizations form the basis for the relationship between the government and religious organizations (Robbers, 2019). This relationship can range from a strict separation, i.e., laicism, resulting in a free religious market (e.g., USA), to a close intermeshing of government and religious organizations, i.e., state churches, leading to a monopolistic religious market dominated by a state church (e.g., Sweden). As with secular markets, the structures of religious markets have important implications for the outcomes generated by these markets. Specifically, Iannaccone (1992, pg. 128) argues that "*pluralistic competition stimulates religious markets just as it does secular* *markets, forcing suppliers to produce efficiently a wide range of alternative faiths well adapted to the specific needs of consumers,*" whereas in countries with a religious monopoly, market forces to achieve the best outcome are lacking.

In this context, Germany ranges in the middle between a free religious market and a monopolistic market, as the Catholic Church and the Protestant Church, officially called the Evangelical Church in Germany, i.e., the two major religious organizations in Germany, split the religious market into a duopoly. Specifically, in 2020, 26.7% of the German population was affiliated with the Catholic Church, and 24.3% was affiliated with the Protestant Church. Furthermore, approximately 3.5% of the population was Muslim, and 4.8% of the population was affiliated with other religious organizations, including minor Christian denominations. Finally, 40.7% of the population was not affiliated with any religious organization and hence does not participate in the religious market.[2]

As in most European countries, religious affiliation in Germany has steadily declined over the past two decades. From 1995 to 2017, the share of members of both the Catholic Church and the Protestant Church decreased from 68.0% to 54.4% of the total population, i.e., 13.5 pp. in total or 0.61 pp. per year. According to a recent projection, the share of the population affiliated with either of the two major religious organizations is expected to further decline to only 31.1% by 2060 (Gutmann and Peters, 2020).

Whereas the Catholic Church in Germany is organized into 27 (arch)dioceses, the Protestant Church is organized into 20 regional churches.[3] In organizational terms, the dioceses and regional churches can be considered independent local "subsidiaries" of the Catholic Church or the Protestant Church managed by a local (arch)bishop, i.e., corresponding to a CEO in secular organizations, with discretionary competences. In addition, individual dioceses and regional churches are characterized by differing management structures and organizational governance systems.

Although the German constitution establishes a separation between religious organizations and the government, there is a constitutionally guaranteed form of cooperation between the government and regional organizations. Specifically, all dioceses and regional churches (and some of the other minor religious organizations) have the status of public corporations, which enables them to exert

2 See `https://fowid.de/meldung/religionszugehoerigkeiten-2020`.

3 The Evangelical Reformed Church as one of the 20 regional churches cannot be attributed to a specific territory, as the church has congregations throughout Germany. Therefore, we only include the remaining 19 (so-called territorial) regional churches.

certain public functions, e.g., establish religious education at public schools, and in turn levy taxes, namely, so-called church taxes, on their members. The church tax is an income tax surcharge of 8-9% (depending on the specific diocese or regional church) of the income tax, which is collected by public fiscal authorities and forwarded to churches (Riegel et al., 2019). Since the public fiscal authorities collect and then forward the church taxes to religious organizations, this marks a stable link between the two churches and the government (Robbers, 2019; Riegel et al., 2018). The link between the two major religious organizations and the government is further strengthened, as both the Catholic and Protestant Church provide a large scale of services within the public welfare system, i.e., kindergartens, schools, hospitals, etc.

Membership in the Catholic Church or the Protestant Church begins with baptism (usually during childhood) and ends either with the death or exit of the church member. Specifically, from the age of 14, individuals can end church membership in front of a public authority (local court, civil registry office, city office or notary), which ends the obligation to pay church taxes (as of next month). In most federal states, the exit requires payment of an administrative fee of up to 60 euros and there might be waiting time for an appointment with the responsible public authority of up to several weeks. In addition to exclusion from religious rites, e.g., the holy communion in the case of the Catholic Church, an exit also excludes the former member from socially and culturally significant events, e.g., obtaining a church marriage, serving as a godfather at baptisms, or receiving a church funeral. Furthermore, the constitutional right of self-determination of religious organizations allows religious organizations to give preferential treatment to their members in employment and service provision in facilities related to the respective religious organization even within the public welfare system. This means that as an indirect effect, for some members, an exit might lead to loss of employment (nonrecruitment) or exclusion from certain public services.

4. Theory and Hypotheses

4.1. Rational Choice of Organizational Affiliation

At its very heart, all organizational research is based on rational choice theory, i.e., the fundamental assumption that individuals make choices that are rational "*within the limits of their information and understanding, restricted by available options, [and] guided by their preferences and tastes*" (Stark and Finke, 2000, p. 65). Accordingly, an individual is expected to always choose from the available options the one that maximizes her individual utility, i.e., the difference between the benefits and costs related to the available options. In the context of organizational affiliation, an individual's decision to affiliate with an organization or disaffiliate from an organization is shaped by the specific benefits and costs of the individual's membership in the organization (Barnard, 1938), i.e., membership benefits and membership costs. Consequently, individuals affiliate with or do not disaffiliate from an organization, and the organization will persist or even grow if they perceive a) their membership benefits to be larger than or at least equal to their membership costs (positive utility of membership) and b) if they perceive their utility of membership in the particular organization to be larger than the utility of membership in a competing organization (Becker, 1976).

In general, membership benefits can be divided into the following three categories: (1) material benefits, (2) solidary benefits, and (3) purposive benefits (Clark and Wilson, 1961). First, material benefits are tangible benefits associated with an individual's membership in an organization, e.g., financial benefits or the access to perks and services reserved for members of an organization. Second, solidary benefits are intangible benefits related to the social aspects of membership in an organization, e.g., provision of a sense of belonging or personal prestige. Finally, purposive benefits are intangible benefits related to the collective pursuit of the organizations' goals, e.g., the spread of the organization's values and the support of consistent social and political goals.

In the context of religious organizations, material membership benefits mainly include (preferential) access to services or employment in facilities related to religious organizations, e.g., nurseries, kindergartens, schools, hospitals, and retirement homes. Solidary benefits include the provision of a personal sense of belonging, social inclusion, or personal prestige, e.g., through participation in socially and culturally significant events, e.g., Sunday service, church marriages, infant baptisms, and church funerals. Purposive benefits typically include achieving illumination or salvation in the afterlife (Azzi and Ehrenberg, 1975; Stark and Bainbridge, 1985), the spread of religious and moral beliefs, and the support of consistent social and political goals through charity work and political work.

Although membership costs are still an underresearched area in the context of organizations (Southby et al., 2019), there is a consensus that due to the limited monetary and time resources of individuals, the monetary costs, e.g., membership fees, and opportunity costs of membership-related time investments, e.g., participation in the activities of the organization, represent the most important cost factors (Southby et al., 2019; Holmes, 2009).

Monetary costs in the context of religious organizations range from (voluntary) contributions to mandatory membership fees and church taxes depending on the respective country-specific regulations and the specific religious organization. Membership-related time investments in Christian religious organizations typically include, inter alia, attendance at Sunday services, confession, or Sunday school. The membership-related time investment amount naturally depends on the specific religious organization and the individual's decision. Specifically, whereas some religious organizations tolerate passive membership, others require active and regular participation.

A scandal related to a particular organization affects both membership benefits and membership costs via at least two channels and, hence, causes members to consider and potentially reevaluate their affiliation with the organization. First, a scandal can reduce solidary benefits due to potential damage to the organization's reputation, which can also negatively affect societal perceptions about the organization's members and, therefore, negatively affect the personal status that results from membership in the scandal-stricken organization. When the organization's reputation is seriously damaged, the solidary benefits might even become negative and, therefore, turn into membership costs. Second, scandals can reduce purposive benefits. Specifically, according to value congruence theory (Chatman, 1989; Edwards and Cable, 2009), membership benefits and membership costs are also affected by the compatibility of values considered important by the affiliated individual and the values conveyed by the organization. Accordingly, congruence between the individual's and the perceived organization's values is an important determinant of an individual's affiliation with an organization.[4] A scandal related to an organization might create a "*discrepancy between [the individual's] expectations and [the] subsequent evaluation of organizational performance*" (Ihm and Baek, 2021, pg. 508). Specifically, a scandal may reduce the congruence between an individual's values and the perceived organization's values and thus result in a decrease in purposive benefits. Moreover, a scandal might even create a divergence between an individual's values and the perceived organization's values (Pfarrer et al., 2010), i.e., create a state of tension resulting in negative purposive benefits.

4 Value congruence is also important for an individual's participation in the activities of an organization. Specifically, Dunaetz et al. (2020) empirically demonstrate a positive relationship between shared values and voluntary work in a church.

4.2. Conceptual Origin of Scandals

Under rational choice theory, not only is an individual's decision to affiliate with or disaffiliate from an organization expected to be formed by rational choices but her behavior within an organization is also. In the context of organizational scandals, an individual engages in misconduct, which might potentially lead to a scandal when the (expected) benefits resulting from the misconduct outweigh the (expected) costs, e.g., punishment, in the event of detection of the misconduct (Draca and Machin, 2015; Lehmann, 2012).

An individual's choice to engage in misconduct is, however, shaped by the organizational environment to a considerable extent. By setting and enforcing rules that govern the conduct of its management and employees, good organizational governance can mitigate organizational misconduct (Rost, 2017; Prakash and Potoski, 2016). Specifically, an organization can reduce the occurrence of opportunities for misconduct and increase the expected costs, e.g., by increasing the detection probability and by holding out the prospect of severe punishment. In turn, this means that poor organizational governance, e.g., missing codes of conduct, nonfunctioning control mechanisms, lack of transparency, (systematic) concealment of misconduct, or a lack of punishment, nourishes misconduct.

Consistent with this, as mentioned in Section 2.1, a strand of prior literature argues that the origin of a scandal, i.e., misconduct, has to be examined both individually and institutionally (Lehmann, 2019; Linstead et al., 2014). We build on this idea and introduce a novel framework for the analysis of scandals that links the conceptual origin of a scandal with its impact on a) the scandal-stricken organization and, due to potential spillover effects of scandals (as discussed in Section 2.3), b) the scandal-stricken organization's competitors. Specifically, our framework (Figure 1), which is based on four quadrants, posits that scandals with different conceptual origins impact the scandal-stricken organization and its competitors differently. We argue that scandals with different conceptual origins also exert different effects on the perceptions of the affiliated individuals and the public's perception of the scandal-stricken organization's values and, therefore, have a different impact on membership benefits and membership costs via solidary and purposive benefits.

The horizontal axis indicates the conceptual origin of a scandal. We discern between a) individual-caused and b) institution-caused scandals . The former group involves scandals resulting from misconduct without any indications that the misconduct was enabled, facilitated or even encouraged by the organization's governance system (quadrants I & III). In contrast, institution-caused scandals result from misconduct directly connected to a poor governance system (quadrants II & IV). Although it is not trivial, especially for outsiders, to determine the role of the organization's governance in a specific case of misconduct and the resulting scandals, there are some publicly observable indicators of institution-caused scandals. Specifically, as there is usually a certain amount of time between the occurrence of a case of misconduct and its publication, the reaction of the organization in the meantime is important for the assessment of the role of the organization's governance. The occurrence of many similar or even connected cases of misconduct, especially over a long period of time, is a strong indicator of an institution-caused scandal, as it implies that the organization failed to set up a code of conduct or efficient control mechanisms to prevent future misconduct. Moreover, if the organization's management has knowledge of the misconduct but does not react adequately, conceals the misconduct, or if the misconduct is grounded in an official organizational policy, this indicates an institution-caused scandal. Furthermore, individual-caused scandals can typically be attributed to a concrete individual or a clearly distinguishable small group of individuals within the organization, whereas institution-caused scandals are often characterized by the involvement of a large group of individuals within the organization. Examples of individual-caused scandals from recent economic history include individual CEO accounting fraud (e.g., Parmalat)[5], or individual CEO insider trading (e.g., Raj Rajaratnam)[6], while examples of institution-caused scandals include organizational accounting fraud (e.g., Enron)[7], and organizational market manipulations (e.g., Deutsche Bank)[8].

The vertical axis finally indicates the impact of a scandal, i.e., the effect of the scandal on the scandal-stricken organization itself (quadrants I & II) and due to potential interorganizational spillover effects the organization's competitors (quadrants III & IV).

[Insert Figure 1 about here]

5 See `https://www.theguardian.com/business/2004/oct/06/corporatefraud.businessqandas`.
6 See `https://www.nytimes.com/2009/11/02/business/02insider.html`.
7 See `https://www.washingtonpost.com/archive/lifestyle/2002/01/30/greedy-liars-the-enron-scandal/bbbe3a86-04d2-4a68-a1b6-758f9ccb6934/`.
8 See `https://www.wsj.com/articles/deutsche-bank-settles-libor-investigation-with-u-s-u-k-authorities-1429791118`.

4.3. Hypothesis Development

We first consider the impact of a scandal on the scandal-stricken organization itself without differentiating by the conceptual origin of the scandal (quadrants I & II of our framework). If we are to understand organizational affiliation on the trajectory of membership benefits and membership costs, we expect a scandal to reduce the utility of membership, i.e., the difference between membership benefits and membership costs, due to its negative impact on the congruence between the individual's and the organization's perceived values reducing the solidary and purposive benefits of membership, as discussed in Section 4.1. However, an individual will only disaffiliate from an organization if the utility of membership turns negative due to the scandal or when the utility declines below the utility of membership in a competing organization. Although monetary benefits that are largely unaffected by scandals might be of some relevance in our setting due to facilities providing public services related to either the Catholic Church or the Protestant Church (see Section 3), they are only of secondary importance compared to solidary and purposive benefits. Specifically, in many church-related facilities, church membership does not represent an exclusive access or employment requirement but rather guarantees preferential treatment. Moreover, in most areas, there are also fully public or private alternatives. A survey among members of the Catholic Church from 2017 indicates that social benefits are even more important reasons for affiliation with the religious organization than purposive benefits. Specifically, a large share of members state that they are members because of family tradition (50%), the desire to baptize their children (50%), the desire to obtain a church funeral (42%), and the desire to obtain a church wedding (42%). Although 51% of members state that they are members because they believe in Jesus Christ, only 39% justify their membership by their belief in the afterlife and the charity work of the church (35%).[9] Membership costs are high in our setting and primarily consist of monetary costs due to the mandatory church tax levied by the Catholic Church and the Protestant Church on their members. In contrast, membership-related time investments can be mostly neglected for Germany's Catholic Church and Protestant Church since passive membership is widespread in both religious organizations. Specifically, in 2019, only

9 See https://de.statista.com/statistik/daten/studie/960349/umfrage/umfrage-unter-katholiken-zu-den-gruenden-fuer-eine-kirchenmitgliedschaft/.

approximately 9.3%[10] of the members of the Catholic Church and 2.7%[11] of the members of the Protestant church regularly participated at Sunday services. Against this background, we argue that scandals have a significant impact on the utility of membership, and we hypothesize the following (in the alternative form):

Hypothesis 1. *Scandals are associated with a decline in affiliation with the scandal-stricken organization.*

In a second step, we separately consider the impacts of scandals with different conceptual origins, i.e., individual-caused (quadrant I of our framework) and institution-caused scandals (quadrant II of our framework), on scandal-stricken organizations because they likely exert different effects on the utility of membership. Specifically, we argue that institution-caused scandals are attributed to the whole organization rather than the individuals who are directly responsible for the misconduct because the misconduct is perceived to be a result of the organization's governance system. Consequently, institution-caused scandals challenge the affiliated individual's and the public's perceptions about the organization's values more than individual-caused scandals and, therefore, exert larger negative effects on solidary and purposive benefits.

On the other hand, individual-caused scandals might involve a personal disappointment of expectations from a specific individual (or a small group of individuals) with whom associated individuals or the public associate the whole organization and, hence, have a stronger negative effect on solidary and purposive benefits than institution-caused scandals. Moreover, because attention to the occurrence of misconduct is a perquisite for the occurrence of a scandal, an attrition effect might be observed, especially for institution-caused scandals. Specifically, as institution-caused scandals are often characterized by the occurrence of many similar cases, media and the public might lose interest in recurring scandals, resulting in weaker effects on solidary and purposive benefits. Therefore, we hypothesize the following (in the alternative form):

Hypothesis 2. *Institution-caused scandals have a different impact on affiliation with the scandal-stricken organization than individual-caused scandals.*

---

10 See `https://de.statista.com/statistik/daten/studie/2640/umfrage/anzahl-von-katholiken-und-katholischen-gottesdienstbesuchern/`.

11 See `https://www.ekd.de/Gottesdienst-Zahlen-Daten-EKD-17289.htm`.

Considering the effects of scandals on the scandal-stricken organization's competitors, we again first consider the effects of individual-caused and institution-caused scandals jointly (quadrants III & IV of our framework). As discussed in Section 2.3, prior research has shown that scandals can exert negative spillover effects, particularly on organizations that are perceived to be somehow associated with the scandal-stricken organization. However, positive spillover effects might be observed for competitors of the scandal-stricken organization not perceived to be associated with the scandal-stricken organization. We do not expect that the religious organizations in our setting experience negative spillover effects following scandals, as they should not be perceived to be associated with each other. Specifically, the Catholic Church and the Protestant Church belong to conceptually distinct categories within Christianity, as Protestantism historically developed in opposition to Catholicism, resulting in a strong categorical boundary between the Protestant Church and the Catholic Church (Piazza and Jourdan, 2018). We therefore hypothesize the following (in the alternative form):

Hypothesis 3. *Scandals are associated with an increase in affiliation with competitors not associated with the scandal-stricken organization.*

In a second step, we consider the effects of individual-caused (quadrant III of our framework) and institution-caused (quadrant IV of our framework) scandals on the organization's competitors separately. In line with our prior proposition that institution-caused scandals have a different effect on affiliation with the scandal-stricken organization compared to individual-caused scandals, we argue that different positive spillover effects on competitors not associated with the scandal-stricken organization can be expected for institution-caused scandals. Therefore, we hypothesize the following (in the alternative form):

Hypothesis 4. *Institution-caused scandals have a different impact on affiliation with competitors not associated with the scandal-stricken organization than individual-caused scandals.*

5. Empirical Strategy

To test H1, we estimate a series of fixed-effects regressions of the following form:

$$C.Affil_{i,t} = \alpha + \beta_1 C.Scands_{i,t} + \beta_2 C.Scands_{i,t-1} + \theta_i + \varphi_t + \varepsilon_{i,t}, \quad (1)$$

where $C.Affil_{i,t}$ is a battery of measures of local affiliation with the Catholic Church in diocese i in year t and $C.Scands_{i,t}$ is the number of scandals in diocese i in year t. We limit our analysis to Catholic dioceses, as most scandals in our setting are related to the Catholic Church. Because the reaction to a scandal might not always be observed immediately, e.g., due to time delays because of the decision-making process, we include $C.Scands_{i,t-1}$, i.e., the number of scandals in diocese i in year t-1. For ease of interpretation, $C.Affil_{i,t}$ is standardized by dividing all values by the sample mean of $C.Affil_{i,t}$ and multiplied by 100 so that the regression coefficients of interest β_1 and β_2 can be interpreted as a percentage of the sample mean. Specifically, $C.Affil_{i,t}$ denotes the following four different measures of affiliation with the Catholic Church in dioceses i and year t: (1) average Sunday service attendance rate ($C.Attendance_{i,t}$), (2) exit rate ($C.Exit_{i,t}$), (3) entry rate ($C.Entry_{i,t}$), and (4) baptism rate ($C.Baptism_{i,t}$), which are all relative to the number of members of the Catholic Church in dioceses i and year t-1. We further include diocese-fixed effects θ_i to control for time-invariant characteristics of individual dioceses that might influence affiliation with the Catholic Church, e.g., socioeconomic factors. Furthermore, we include year-fixed effects φ_t to control for time-variant national trends in affiliation resulting from nationwide scandals or non-scandal events, e.g., the election of a new Pope, and intraorganizational spillover effects of local scandals, i.e., the effect of local scandals on affiliation beyond the borders of the diocese where the scandal occurred. However, in the case of intraorganizational spillover effects of local scandals, including φ_t into the regression would impose a downward bias on our results, as β_1 and β_2 capture only the additional local effect of a scandal that exceeds the nationwide effect of the scandal. Consequently, our results represent the lower bound of the effects of scandals on affiliation. We therefore additionally estimate an equation without φ_t for all our analyses. H1 predicts a decline in affiliation with the scandal-stricken organization following a scandal. Consequently, H1 is confirmed if either β_1 or β_2 or both are significantly negative, apart from regressions with $C.Exit_{i,t}$ as the independent variable where either a significantly positive β_1 or β_2 or both support H1.

To test H2, we estimate a series of fixed-effects regressions with measures of local affiliation with the Catholic Church as dependent variables and separate independent variables for present and lagged local individual-caused and institution-caused Catholic Church scandals. Specifically, we estimate regressions of the following form:

$$C.Affil_{i,t} = \alpha + \beta_1 C.ins.Scands_{i,t} + \beta_2 C.ins.Scands_{i,t-1} + \beta_3 C.idv.Scands_{i,t} + \beta_4 C.idv.Scands_{i,t-1} + \theta_i + \varphi_t + \varepsilon_{i,t}, \tag{2}$$

where $C.ins.Scands_{i,t(-1)}$ is the present (lagged) number of institution-caused scandals and $C.idv.Scands_{i,t(-1)}$ is the present (lagged) number of individual-caused scandals in diocese i and year t(-1). H2 predicts a different effect of institution-caused scandals on affiliation with the scandal-stricken organization compared to individual-caused scandals. Consequently, H2 is confirmed if coefficients β_1 and β_2 are different from coefficients β_3 and β_4.

To test H3, we estimate a series of fixed-effects regressions with measures of local affiliation with the Protestant Church as the dependent variable and the number of present and lagged local Catholic scandals as the independent variables of interest. Accordingly, we estimate regressions of the following form:

$$P.Affil_{j,t} = \alpha + \beta_1 C.Scands_{j,t} + \beta_2 C.Scands_{j,t-1} + P.Scands_{j,t} + P.Scands_{j,t-1} + \theta_j + \varphi_t + \varepsilon_{j,t}, \tag{3}$$

where $P.Affil_{j,t}$ are four measures of local affiliation with the Protestant Church in regional church j in year t, which correspond to the measures of affiliation with the Catholic Church introduced above, i.e., $P.Attendance_{j,t}$, $P.Exit_{j,}$, $P.Entriy_{j,t}$, and $P.Baptism_{j,t}$. $C.Scands_{j,t(-1)}$ is the present (lagged) number of Catholic scandals in regional church j and year t(-1). Because there is a considerable number of coincident Catholic and Protestant scandals, we also include the present (lagged) number of Protestant scandals $P.Scands_{j,t(-1)}$ in regional church j and year t(-1) as a control variable. H3 predicts that scandals are associated with an increase in affiliation with competitors not affiliated with the scandal-stricken organization. Therefore, H3 is confirmed if either β_1 or β_2 or both are significantly positive. For regressions with $P.Exit_{j,t}$ as the independent variable, no significant coefficients β_1 and β_2 should be expected.

Finally, to test H4, we estimate a series of fixed-effects regressions with measures of local affiliation with the Protestant Church as dependent variables and separate independent variables for present and lagged local individual-caused and institution-caused Catholic Church scandals. Specifically, we estimate regressions of the following form:

$$P.Affil_{j,t} = \alpha + \beta_1 C.ins.Scands_{j,t} + \beta_2 C.ins.Scands_{j,t-1} + \beta_3 C.idv.Scands_{j,t} + \beta_4 C.idv.Scands_{j,t-1} + P.Scands_{j,t} + P.Scands_{j,t-1} + \theta_j + \varphi_t + \varepsilon_{j,t}, \tag{4}$$

where $C.ins.Scands_{j,t(-1)}$ is the present (lagged) number of institution-caused scandals and $C.idv.Scands_{j,t(-1)}$ is the present (lagged) number of individual-caused scandals in regional church j and year t(-1). H4 predicts a different effect of institution-caused scandals on affiliation with the scandal-stricken organization's competitors compared to individual-caused scandals. Consequently, H4 is confirmed if coefficients β_1 and β_2 are different from coefficients β_3 and β_4.

6. Data

6.1. Local Catholic Church and Protestant Church Scandals

To obtain data on local Catholic Church scandals ($C.Scands_{i,t}$) and Protestant Church scandals ($P.Scands_{j,t}$), i.e., scandals related to a single diocese or regional church, we search German media publications, e.g., newspapers, magazines, web-based publications, and newswires, using the LexisNexis databank. We restrict our search to scandals that can be clearly attributed to a particular diocese or regional church rather than nationwide scandals, e.g., a controversial decision of the Pope as the supreme authority of the Catholic Church. We identify 121 local scandals involving the Catholic Church and 16 local scandals involving the Protestant Church between 2002 and 2016. The scandals can be divided into the following six broad categories: financial scandals, scandals related to the church as an employer, scandals related to the church as a public service provider, scandals related to child pornography, scandals related to sexual abuse, and other scandals (e.g., scandals related to homophobic statements by church officials). We then classify the scandals by their conceptual origin according to the publicly observable indicators described in Section 4.2. Specifically, we assess whether a scandal is institution-caused or individual-caused by considering the occurrence of similar or even connected cases of misconduct, information about the organization's management knowledge of the misconduct, concealment of the misconduct by the organization, or whether the misconduct is somehow grounded in an official organizational policy, e.g., an official guideline that allows the release of employees after they obtain a divorce.

We report the distribution of the scandals included in our sample by time, category, and conceptual origin in Table 1. Panel A reports Catholic Church scandals, and Panel B reports Protestant

Church scandals.

[Insert Table 1 about here]

Most scandals involving the Catholic Church (56) and the Protestant Church (9) occurred in 2010, and most Catholic Church scandals (91) and Protestant Church scandals (11) were related to sexual abuse. Moreover, the vast majority of scandals are institution-caused scandals for the Catholic Church (111) and the Protestant Church (12).

The geographic distribution of Catholic Church scandals, which represent almost 90% of all scandals included in our sample over the sample period, is reported in Figure 2.

[Insert Figure 2 about here]

The Catholic Church scandals are geographically dispersed across most of the German dioceses. In fact, there were only two dioceses with zero scandals between 2002 and 2016. However, the dioceses in southern Germany and western Germany, which are characterized by a relatively large share of Catholic population, faced substantially more scandals than the dioceses in eastern and northern Germany, which are characterized by a relatively low share of Catholic population.

6.2. Measures of Local Affiliation with the Catholic Church and the Protestant Church

To obtain data on local affiliation with the Catholic Church ($C.Affil_{i,t}$) and the Protestant Church ($P.Affil_{j,t}$), we hand-collect genuine data from annual reports made available by the two churches. Specifically, as discussed in Section 5, we collect data on four different measures, i.e., (1) average Sunday service attendance rate, (2) exit rate, (3) entry rate, and (4) baptism rate at the level of Germany's 27 Catholic dioceses and Germany's 19 Protestant regional churches for the years 2002-2016.[12]

The measures of affiliation differ substantially in regard to the time delay between the occurrence of a scandal and the realization of disaffiliation. Specifically, an individual can stop attending Sunday service immediately, but as discussed in Section 3, it may take some time to obtain an appointment to officially declare the exit in front of a public authority. Along the same lines, entry into a church

12 We collect data on the number of local members for the years 2001-2016, as all measures of affiliation are relative to the number of members in year t-1.

might be delayed. Refraining from the baptism of children might take even more time to be realized, as such events are usually planned over a longer horizon. Moreover, the measures also differ in the degree of (dis)affiliation, i.e., the reversibility and long-term consequences. Whereas the decision not to attend Sunday service can be reversed easily, reversing the exit or entry from church requires more effort. Although theoretically possible, refraining from baptizing children is practically irreversible and hence indicates a high degree of disaffiliation.

As individual dioceses and regional churches differ by their size, we compute all measures relative to the number of members of the Catholic Church or the Protestant Church in dioceses i or regional church j in year t-1 to achieve comparability. For ease of interpretation, both $C.Affil_{i,t}$ and $P.Affil_{i,t}$ are standardized by dividing all values by the sample mean of $C.Affil_{i,t}$ and $P.Affil_{i,t}$, respectively, and multiplied by 100 so that the regression coefficients of interest can be interpreted as the percentage deviation from the sample mean.

Descriptive statistics of raw, i.e., nonstandardized, measures of affiliation $C.Affil_{i,t}$ ($rC.Affil_{i,t}$, Panel A) and $P.Affil_{i,t}$ ($rP.Affil_{i,t}$, Panel B) are reported in Table 2. We report nonstandardized values here to allow a better understanding of the data. To provide a first indication of the effect of scandals on affiliation, we report descriptive statistics separately for the dioceses and regional churches that faced at least one scandal, i.e., that are scandal-stricken, and those that did not face any scandal over the sample period.

[Insert Table 2 about here]

Catholic dioceses are on average smaller than Protestant regional churches in terms of total population (6,874,307 vs. 3,487,323) and the number of church members (2,030,769 vs. 1,370,767). However, on average, the Catholic dioceses have a higher percentage of members in the total population (43.5% vs. 32.31%). Regarding the measures of affiliation, Catholic dioceses are on average characterized by a substantially higher Sunday service attendance rate (12.79% vs. 3.05%). On the other hand, Protestant regional churches have on average a higher entry rate (0.14% vs. 0.01%) and baptism rate (0.83% vs. 0.72%). However, the exit rate is on average also higher for Protestant regional churches (0.67% vs. 0.54%).

Scandal-stricken Catholic dioceses face on average 1.67 scandals over the sample period. Whereas most measures of affiliation are similar in scandal-stricken and nonstricken dioceses, the exit rate is substantially higher in scandal-stricken dioceses (0.64% vs. 0.51%). A similar pattern is observable for scandal-stricken Protestant regional churches, which face on average 1.43 scandals and which have a substantially higher exit rate in scandal-stricken regional churches compared to nonstricken regional churches (0.77% vs. 0.66%).

7. Empirical Evidence

7.1. Effects of Scandals on Scandal-Stricken Organizations

7.1.1. Joint Effect of Individual- and Institution-Caused Scandals

We first provide a naive graphical assessment of the effects of scandals on affiliation with the Catholic Church. Figure 3 shows the raw exit rates ($rC.Exit_{i,t}$) of scandal-stricken Catholic dioceses ($C.Scands_{i,t} > 0$) in the respective year (circle markers) and the nonstricken dioceses ($C.Scands_{i,t} = 0$) in the respective year (x markers) for the years 2002 and 2016. The red (black) dashed line indicates the annual mean $rC.Exit_{i,t}$ for scandal-stricken (nonstricken) dioceses weighted by the one-year lagged number of members.

[Insert Figure 3 about here]

There is considerable variation in $rC.Exit_{i,t}$ across dioceses and years. However, in general, $rC.Exit_{i,t}$ is higher in scandal-stricken dioceses than in nonstricken dioceses. This provides the first indication that scandals are negatively associated with affiliation with the scandal-stricken organization.

Table 3 reports the regression results from estimating Equation 1 for the four measures of affiliation with the Catholic Church ($C.Affill_{i,t}$).

[Insert Table 3 about here]

Considering our preferred specification, which includes diocese-fixed effects (Columns 1b-4b), both the present and the lagged coefficient for *C.Attendance* are insignificant. Accordingly, scandals seem not to have any impact on the Sunday service attendance rate, which is our measure of affiliation characterized by the lowest time delay and lowest degree of (dis)affiliation. However, as the overall attendance rate is relatively low, i.e., only approximately 12% of the members of the Catholic Church attend Sunday service, *C.Attendance* should be taken with care as a measure of affiliation because it might capture only the most loyal members, i.e., those with the highest membership benefits. These are hence the last to disaffiliate from the organization after a scandal. The present coefficient for *C.Exit* is significantly positive and indicates that one local scandal increases the local exit rate by almost 4.1% of the sample mean in the year of the scandal. As the coefficient is no longer significant when considering the one-year lag of the scandal, this provides some indication that scandals exert only a short-term effect on the exit rate from the scandal-stricken organization. Both coefficients are insignificant for *C.Entry*, which is not surprising, as there is no reason to expect that a scandal would increase the entry rate of the scandal-stricken organization. Finally, the lagged coefficient for *C.Baptism* is significantly positive, indicating that one scandal reduces the baptism rate by 0.67% of the sample mean in the year following the scandal. Recall, that this measure is characterized by the highest time delay and the highest degree of disaffiliation. Considered jointly, the results are consistent with H1, which predicts a negative effect of scandals on affiliation with the scandal-stricken organization.

7.1.2. Separate Effects of Individual- and Institution-Caused Scandals

Table 4 reports the regression results from estimating Equation 2 for the four measures of affiliation with the Catholic Church separately for individual-caused and institution-caused scandals.

[Insert Table 4 about here]

Considering our preferred specification, the results indicate that both individual-caused and institution-caused scandals increase the exit rate (*C.Exit*) from the Catholic Church in the year of the scandal. However, whereas an individual-caused local scandal leads to an increase in the local exit rate of 15.6% of the sample mean, an institution-caused scandal increases the exit rate by only 3.5% of the sample mean. Similarly, an individual-caused scandal reduces the baptism rate (*C.Baptism*) in the year after the scandal by 1.94% of the sample mean, whereas an institution-caused scandal reduces the baptism rate by only 0.61% of the sample mean. Again, neither institution-caused nor individual-caused scandals affect the entry rate (*C.Entry*). Surprisingly, individual-caused scandals increase the attendance rate (*C.Attendance*) in the year of the scandal and in the year after the scandal by 2.1% and 3.8% of the sample mean, respectively. This is consistent with the argument that *C.Attendance* must be taken with care as a measure of affiliation. Taken together, the results are consistent with H2, which predicts a different effect of institution-caused scandals on affiliation with the scandal-stricken organization than individual-caused scandals. Specifically, individual-caused scandals seem to have a stronger negative effect on affiliation than institution-caused scandals. The stronger effects of individual-caused scandals might be either the result of a personal disappointment regarding expectations of a particular individual (or a small group of individuals) with whom affiliated individuals or the public associate the whole organization or an attrition effect of institution-caused scandals that are often characterized by the occurrence of many similar cases which can result in a loss of interest in the scandal by the media and public.

7.2. Effects of Scandals on Scandal-Stricken Organizations' Competitors

7.2.1. Joint Effect of Individual- and Institution-Caused Scandals

We again first provide a graphical assessment of the effects of scandals on the scandal-stricken organization's unassociated competitors. Figure 4 shows the raw exit rates ($rP.Exit_{j,t}$) of Protestant regional churches exposed to only Catholic Church scandals ($C.Scands_{j,t} > 0$ & $P.Scands_{j,t} = 0$) (circle markers), regional churches exposed to only Protestant Church scandals ($P.Scandsj, t > 0$ & $C.Scands_{j,t} = 0$) (square markers), regional churches exposed to both Protestant Church scandals and Catholic Church scandals ($C.Scands_{j,t} > 0$ & $P.Scands_{j,t} > 0$) (triangle markers), and regional churches not exposed to any scandal ($C.Scands_{j,t} = 0$ & $P.Scands_{j,t} = 0$) (x markers) in the respective year for the years 2002 and 2016. The black dashed line indicates the annual mean $rP.Exit_{j,t}$ for regional churches not exposed to any scandal weighted by the one-year lagged number of members.

[Insert Figure 4 about here]

As with the exit rates of Catholic dioceses, there is large variation in exit rates across regional churches and years. Furthermore, exit rates in regional churches exposed to only a Protestant Church scandal or both a Protestant Church scandal and a Catholic Church scandal seem to be higher than average exit rates of regional churches not exposed to any scandal. Moreover, exit rates of regional churches exposed to a Catholic Church scandal only do not exceed the average exit rate of regional churches not exposed to any scandal. This provides a first indication of an absence of negative interorganizational spillover effects of scandals.

Table 5 reports the regression results from estimating Equation 3 for the four measures of affiliation with the Protestant Church ($P.Affil_{i,t}$).

[Insert Table 5 about here]

Considering our preferred specification, which includes regional church-fixed effects (Columns 1b-4b), the results indicate that Catholic Church scandals have no effect on the exit rate (*P.Exit*) and attendance rate (*P.Attendance*) of the Protestant Church in the year of the scandal or the following year. However, a Catholic Church scandal, increases the entry rate (*P.Entry*) of the Protestant Church by 1.6% of the sample mean in the year of the scandal and by 0.9% of the sample mean in the following year. There is also a negative effect of Catholic scandals on the baptism rate of the Protestant Church (*P.Baptism*) in the year following the scandal of 0.3% of the sample mean. However, the effect is only weakly significant and small in economic terms.[13] Taken together, the results are consistent with H3, which predicts positive spillover effects of scandals on affiliation with unassociated competitors of the scandal-stricken organization.

7.2.2. Separate Effects of Individual- and Institution-Caused Scandals

Table 6 reports the regression results from estimating Equation 4 for the four measures of affiliation with the Protestant Church separately for individual-caused and institution-caused scandals.

[Insert Table 6 about here]

Considering our preferred specification, the results indicate that both institution-caused and individual-caused Catholic Church scandals have no effect on the exit rate from the Protestant Church (*P.Exit*). Moreover, individual-caused Catholic Church scandals have a weakly significant negative effect on the attendance rate of the Protestant Church (*P.Attendance*) in the year of the scandal, and institution-caused Catholic Church scandals have a weakly significant negative effect on the baptism rate of the Protestant Church (*P.Baptism*) in the year following the scandal. However, only institution-caused Catholic Church scandals have a significantly positive effect on the entry rate of the Protestant Church (*P.Entry*) both in the yar of the scandal and in the following year. Specifically, an institution-caused Catholic Church scandal increases the entry rate of the Protestant Church by 1.6% (1.3%) of the sample mean in the year of the scandal (in the following year). Considered jointly, the results are consistent with H3, which predicts different effects of institution-caused

13 There are no significant effects of Protestant Church scandals on affiliation with the Protestant Church, which can likely be attributed to the low number of Protestant scandals in our sample.

and individual-caused scandals on unassociated competitors of the scandal-stricken organization. Specifically, although there is some noise regarding the baptism rate and the attendance rate, the results suggest that institution-caused scandals have a stronger positive effect on affiliation with the scandal-stricken organization's competitors. This is surprising at first glance, as individual-caused scandals seem to have a stronger negative effect on affiliation with the scandal-stricken organization, and we hence would expect stronger positive spillover effects of individual-caused scandals on unassociated competitors. A potential explanation might be that institution-caused and individual-caused scandals cause reactions from different types of members of the scandal-stricken organization. Specifically, individual-caused scandals might result in disaffiliation of individuals with on average lower benefits of membership in religious organizations in general (e.g., less religious individuals) compared to individuals who disaffiliate after institution-caused scandals. Such individuals might be less likely to join a competing religious organization after disaffiliating from the scandal-stricken organization.

8. Conclusion

This paper adds to the emerging field of research on the effects of scandals on organizations and their stakeholders. We introduce a new framework for the analysis of scandals that links the conceptual origin of a scandal, i.e., individual-caused vs. institution-caused, with its impact on the scandal-stricken organization vs. the scandal-stricken organization's competitors. We analyze the effects of almost 140 locally reported scandals that occurred in Germany's 27 Catholic dioceses and 19 Protestant regional churches between 2002 and 2016 to examine the effects of scandals on organizational affiliation and competition. Specifically, using genuine diocese-level and regional church-level data on affiliation with the two major religious organizations in Germany, we study (1) the association between scandals and affiliation with the scandal-stricken organization, (2) whether individual-caused vs. institution-caused scandals have different effects on affiliation with the scandal-stricken organization, (3) whether scandals have interorganizational spillover effects on the the scandal-stricken organization's competitors, and (4) whether these spillover effects differ for individual-caused and institution-caused scandals.

We find that both individual-caused and institution-caused scandals are associated with a decline in affiliation with the scandal-stricken organization. However, individual-caused scandals have a significantly larger effect on affiliation with the scandal-stricken organization than institution-caused scandals. We also find evidence for positive interorganizational spillover effects on unassociated competitors of the scandal-stricken organization but only for institution-caused scandals. Our results contribute to the research on religious disaffiliation and to the emerging field of studies on the effects of scandals on organizations and their stakeholders. We argue that due to the economic character of religious organizations, our results can be generalized beyond our empirical setting, particularly to stakeholders of secular organizations.

References

Adut, A. (2008). *On scandal: Moral disturbances in society, politics, and art*. Cambridge University Press New York.

Azzi, C. and Ehrenberg, R. (1975). Household Allocation of Time and Church Attendance. *Journal of Political Economy*, 83(1):27–56.

Barnard, C. I. (1938). The economy of incentives. *Classics of organization theory*, pages 93–102.

Becker, G. S. (1976). *The economic approach to human behavior*. University of Chicago press.

Bottan, N. L. and Perez-Truglia, R. (2015). Losing my religion: The effects of religious scandals on religious participation and charitable giving. *Journal of Public Economics*, 129:106–119.

Bouzzine, Y. D. and Lueg, R. (2020). The contagion effect of environmental violations: The case of Dieselgate in Germany. *Business Strategy and the Environment*, 29(8):3187–3202.

Bundy, J. and Pfarrer, M. D. (2015). A burden of responsibility: The role of social approval at the onset of a crisis. *Academy of Management Review*, 40(3):345–369.

Byun, K.-A., Duhan, D. F., and Dass, M. (2020). The preservation of loyalty halo effects: An investigation of the postproduct-recall behavior of loyal customers. *Journal of Business Research*, 116(2):163–175.

Chatman, J. A. (1989). Improving Interactional Organizational Research: A Model of Person-Organization Fit. *Academy of Management Review*, 14(3):333–349.

Clark, P. B. and Wilson, J. Q. (1961). Incentive Systems: A Theory of Organizations. *Administrative Science Quarterly*, 6(2):129–166.

Clemente, M. and Gabbioneta, C. (2017). How Does the Media Frame Corporate Scandals? The Case of German Newspapers and the Volkswagen Diesel Scandal. *Journal of Management Inquiry*, 26(3):287–302.

Coombs, W. T. (1995). Choosing the right words: The development of guidelines for the selection of the "appropriate" crisis-response strategies. *Management Communication Quarterly*, 8(4):447–476.

Del Triana, M. C., Jayasinghe, M., Pieper, J. R., Delgado, D. M., and Mingxiang, L. (2019). Perceived Workplace Gender Discrimination and Employee Consequences: A Meta-Analysis and Complementary Studies Considering Country Context. *Journal of Management*, 45(6):2419–2447.

Draca, M. and Machin, S. (2015). Crime and economic incentives. *Annual Review of Economics*, 7(1):389–408.

Dunaetz, D. R., Smyly, C., Fairley, C., and Heykoop, C. (2020). Values congruence and organizational commitment in churches: When do shared values matter? *Psychology of Religion and Spirituality. Advance online publication.*

Edwards, J. R. and Cable, D. M. (2009). The value of value congruence. *Journal of applied psychology*, 94(3):654.

Ewelt-Knauer, C., Knauer, T., and Lachmann, M. (2015). Fraud characteristics and their effects on shareholder wealth. *Journal of Business Economics*, 85(9):1011–1047.

Frick, B. and Simmons, R. (2017). The impact of exogenous shocks on exits from the Catholic and Protestant churches in Germany, 1953–2015. *Applied Economics Letters*, 24(20):1476–1480.

Gadgil, S. and Sockin, J. (2020). Caught in the Act: How Corporate Scandals Hurt Employees.

Goldman, B. M., Gutek, B. A., Stein, J. H., and Lewis, K. (2006). Employment Discrimination in Organizations: Antecedents and Consequences. *Journal of Management*, 32(6):786–830.

Gutmann, D. and Peters, F. (2020). German Churches in Times of Demographic Change and Declining Affiliation: A Projection to 2060. *Comparative Population Studies*, 45:4–34.

Hersel, M. C., Helmuth, C. A., Zorn, M. L., Shropshire, C., and Ridge, J. W. (2019). The corrective actions organizations pursue following misconduct: A review and research agenda. *Academy of Management Annals*, 13(2):547–585.

Holmes, K. (2009). The value of volunteering: The volunteer's story. *Australian Journal on Volunteering*, 14:50–58.

Hungerman, D. M. (2013). Substitution and stigma: Evidence on religious markets from the catholic sex abuse scandal. *American Economic Journal: Economic Policy*, 5(3):227–253.

Iannaccone, L. R. (1992). Religious markets and the economics of religion. *Social compass*, 39(1):123–131.

Ihm, J. and Baek, Y. M. (2021). Why do participants in voluntary organizations leave? Exploring the relationship between value congruence and length of stay. *Nonprofit Management and Leadership*, 31(3):505–524.

Jonsson, S., Greve, H. R., and Fujiwara-Greve, T. (2009). Undeserved loss: The spread of legitimacy loss to innocent organizations in response to reported corporate deviance. *Administrative Science Quarterly*, 54(2):195–228.

Karpoff, J. M., Lee, D. S., and Vendrzyk, V. P. (1999). Defense procurement fraud, penalties, and contractor influence. *Journal of Political Economy*, 107(4):809–842.

Knittel, C. R. and Stango, V. (2014). Celebrity Endorsements, Firm Value, and Reputation Risk: Evidence from the Tiger Woods Scandal. *Management Science*, 60(1):21–37.

Korkofingas, C. and Ang, L. (2011). Product recall, brand equity, and future choice. *Journal of Marketing Management*, 27(9-10):959–975.

Lange, D. and Washburn, N. T. (2012). Understanding attributions of corporate social irresponsibility. *Academy of Management Review*, 37(2):300–326.

Lehmann, E. E. (2012). Corporate Governance, Compliance & Crime. In Rotsch, T., editor, *Wissenschaftliche und praktische Aspekte der nationalen und internationalen Compliance-Diskussion*, pages 53–72. Nomos Verlagsgesellschaft mbH & Co. KG.

Lehmann, E. E. (2019). Corporate Governance in Business and Management. In *Oxford Research Encyclopedia of Business and Management*.

Linstead, S., Maréchal, G., and Griffin, R. W. (2014). Theorizing and Researching the Dark Side of Organization. *Organization Studies*, 35(2):165–188.

Madera, J. M., King, E. B., and Hebl, M. R. (2012). Bringing social identity to work: The influence of manifestation and suppression on perceived discrimination, job satisfaction, and turnover intentions. *Cultural Diversity and Ethnic Minority Psychology*, 18(2):165.

Monga, A. and John, D. (2008). When does negative brand publicity hurt? The moderating influence of analytic versus holistic thinking. *Journal of Consumer Psychology*, 18(4):320–332.

Ni, J., Flynn, B. B., and Jacobs, F. R. (2016). The effect of a toy industry product recall announcement on shareholder wealth. *International Journal of Production Research*, 54(18):5404–5415.

Paruchuri, S. and Misangyi, V. F. (2015). Investor perceptions of financial misconduct: The heterogeneous contamination of bystander firms. *Academy of Management Journal*, 58(1):169–194.

Pfarrer, M. D., Pollock, T. G., and Rindova, V. P. (2010). A tale of two assets: The effects of firm reputation and celebrity on earnings surprises and investors' reactions. *Academy of Management Journal*, 53(5):1131–1152.

Piazza, A. and Jourdan, J. (2018). When the dust settles: The consequences of scandals for organizational competition. *Academy of Management Journal*, 61(1):165–190.

Prakash, A. and Potoski, M. (2016). Dysfunctional institutions? Toward a new agenda in governance studies. *Regulation & Governance*, 10(2):115–125.

Rhee, M. and Haunschild, P. R. (2006). The Liability of Good Reputation: A Study of Product Recalls in the U.S. Automobile Industry. *Organization Science*, 17(1):101–117.

Riegel, U., Gutmann, D., Peters, F., and Faix, T. (2019). Does church tax matter?: The influence of church tax on leaving the church. *International Journal of Practical Theology*, 23(2):168–187.

Riegel, U., Kröck, T., and Faix, T. (2018). Warum Menschen die katholische Kirche verlassen. Eine explorative Untersuchung zu Austrittsmotiven im Mixed-Methods-Design. In Etscheid-Stams, M., Laudage-Kleeberg, R., and Rünker, T., editors, *Kirchenaustritt - oder nicht?*, pages 125–2017. Herder, Freiburg and Basel and Wien.

Robbers, G. (2019). *State and church in the European Union*. Nomos Verlag.

Röck, T., Bramkamp, M., Bartz-Schmidt, K. U., and Röck, D. (2017). Organ transplantation scandal influencing corneal donation rate. *International Journal of Ophthalmology*, 10(6):1001–1003.

Rost, K. (2017). Introduction to the corporate governance of religion. In *Emerging trends in the social and behavioral sciences*, pages 1–15. John Wiley & Sons, Ltd.

Song, C. and Han, S. H. (2017). Stock Market Reaction to Corporate Crime: Evidence from South Korea. *Journal of Business ethics*, 143(2):323–351.

Southby, K., South, J., and Bagnall, A.-M. (2019). A Rapid Review of Barriers to Volunteering for Potentially Disadvantaged Groups and Implications for Health Inequalities. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 30(5):907–920.

Stark, R. and Bainbridge, W. S. (1985). 19. Secularization, Revival, and Cult Formation. In Stark, R. and Bainbridge, W. S., editors, *The Future of Religion*, pages 429–456. University of California Press.

Stark, R. and Finke, R. (2000). Catholic Religious Vocations: Decline and Revival. *Review of Religious Research*, 42(2):125.

Tarrow, S. (1994). Power in movement: Collective action, social movements and politics.

Thompson, J. B. (2013). *Political scandal: Power and visability in the media age.* John Wiley & Sons.

Tibbs, S. L., Harrell, D. L., and Shrieves, R. E. (2011). Do Shareholders Benefit from Corporate Misconduct? A Long-Run Analysis. *Journal of Empirical Legal Studies*, 8(3):449–476.

Wang, S. and Alexander, P. (2018). The factors of consumer confidence recovery after scandals in food supply chain safety. *Asia Pacific Journal of Marketing and Logistics*, 30(5):1379–1400.

Wooten, L. P. and James, E. H. (2008). Linking crisis management and Lleadership competencies: The role of human resource development. *Advances in Developing Human Resources*, 10(3):352–379.

Zavyalova, A., Pfarrer, M. D., Reger, R. K., and Hubbard, T. D. (2016). Reputation as a benefit and a burden? How stakeholders' organizational identification affects the role of reputation following a negative event. *Academy of Management Journal*, 59(1):253–276.

Zhou, Y. and Makridis, C. (2019). Financial Misconduct and Changes in Employee Satisfaction.

Figure 1: Framework for the Analysis of Scandals by their Conceptual Origin

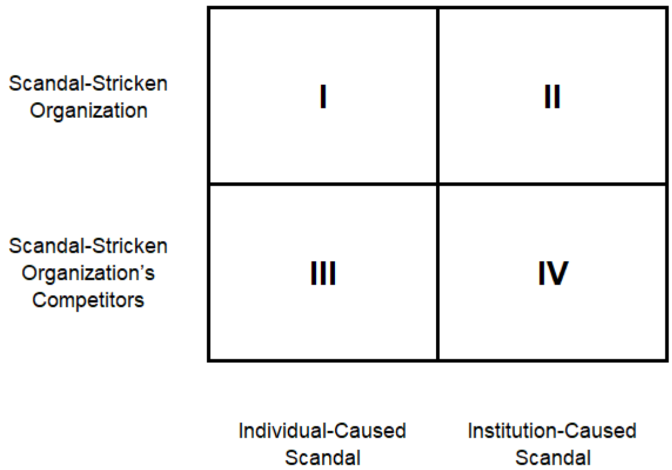


Figure 2: Geographic Distribution of Catholic Church Scandals, 2002-2016

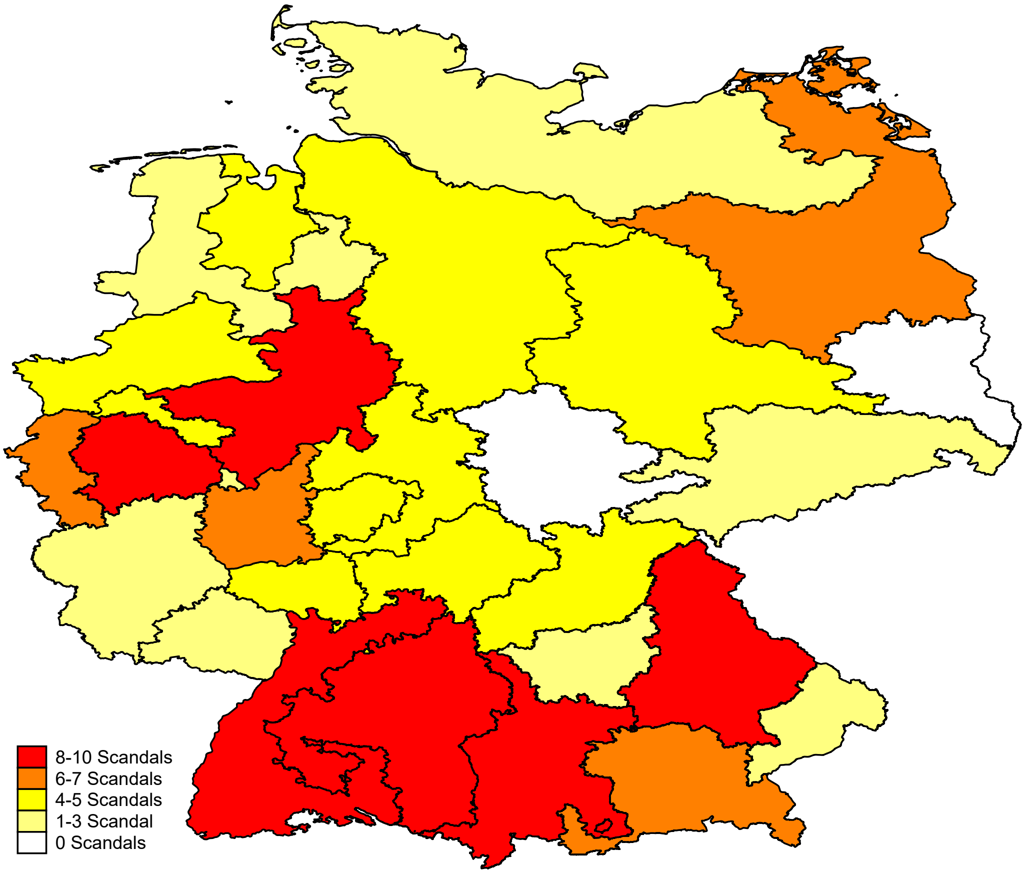


Figure 3: Raw Exit Rates of the Catholic Church, 2002-2016

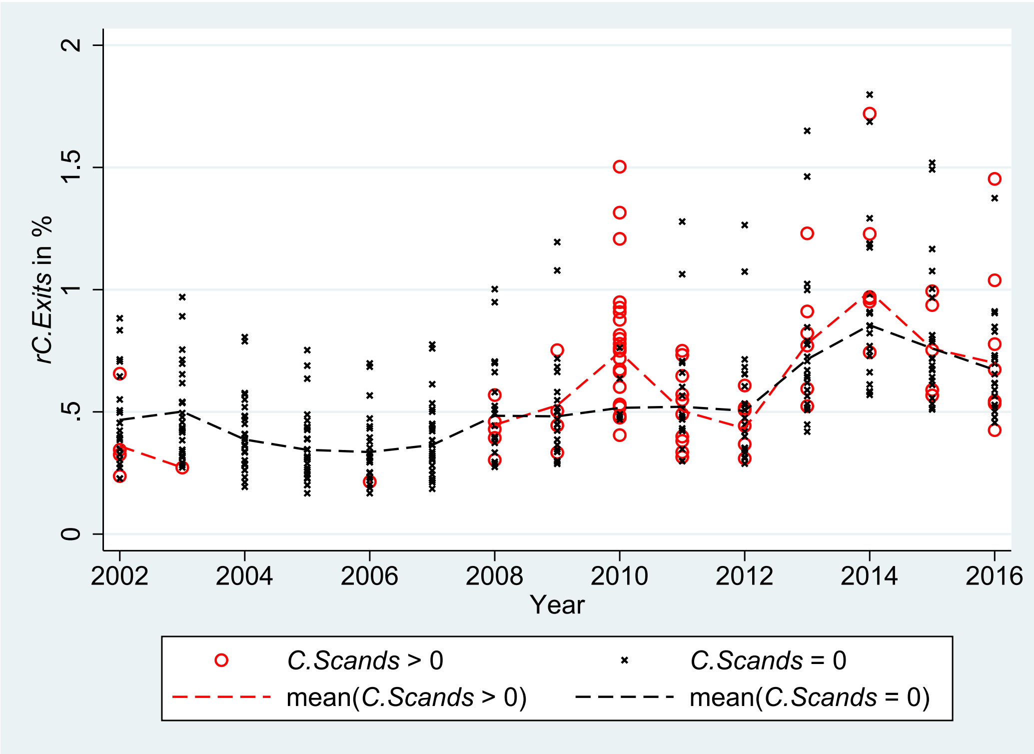


Figure 4: Raw Exit Rates of the Protestant Church, 2002-2016

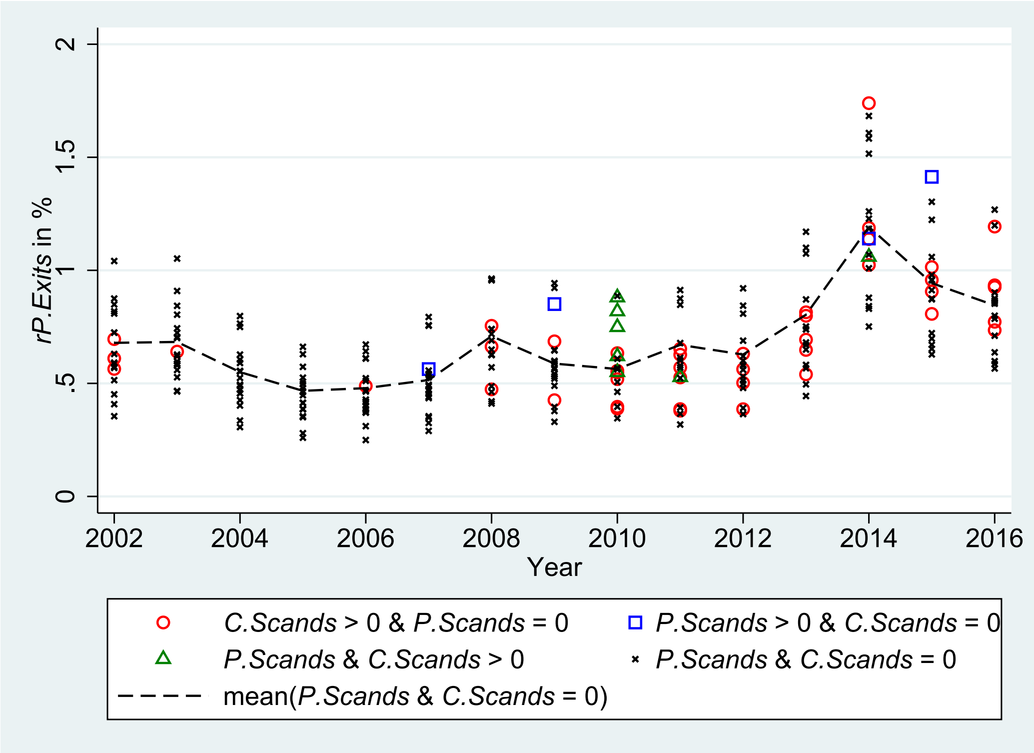


Table 1: Descriptive Statistics: Catholic Church Scandals and Protestant Church Scandals, 2002-2016

Panel A: Catholic Church Scandals

Year	C.Scandals	Conceptual Origin		Category of Scandals					
		C.idv.Scands	C.ins.Scands	Financial	Employer	Public Service	Pornography	Sexual Abuse	Other
2002	5	0	5	0	0	0	0	5	0
2003	1	0	1	0	0	0	0	1	0
2004	0	0	0	0	0	0	0	0	0
2005	0	0	0	0	0	0	0	0	0
2006	1	1	0	1	0	0	0	0	0
2007	0	0	0	0	0	0	0	0	0
2008	6	0	6	0	0	0	1	5	0
2009	4	1	3	1	0	0	1	2	0
2010	56	3	53	2	1	1	0	51	1
2011	12	1	11	1	1	0	2	8	0
2012	7	0	7	0	2	0	2	3	0
2013	7	1	6	1	0	1	1	4	0
2014	9	1	8	1	1	0	3	4	0
2015	6	1	5	0	0	1	0	4	1
2016	7	1	6	0	1	0	1	4	1
Total	**121**	**10**	**111**	**7**	**6**	**3**	**11**	**91**	**3**

Panel B: Protestant Church Scandals

Year	P.Scands	Conceptual Origin		Category of Scandals					
		P.idv.Scands	P.ins.Scands	Financial	Employer	Public Service	Pornography	Sexual Abuse	Other
2002	0	0	0	0	0	0	0	0	0
2003	0	0	0	0	0	0	0	0	0
2004	0	0	0	0	0	0	0	0	0
2005	0	0	0	0	0	0	0	0	0
2006	0	0	0	0	0	0	0	0	0
2007	1	0	1	0	0	0	0	1	0
2008	0	0	0	0	0	0	0	0	0
2009	1	0	1	0	0	0	1	0	0
2010	9	1	8	0	0	0	0	8	1
2011	1	1	0	1	0	0	0	0	0
2012	0	0	0	0	0	0	0	0	0
2013	0	0	0	0	0	0	0	0	0
2014	3	1	2	1	0	0	0	2	0
2015	1	1	0	0	0	0	0	0	1
2016	0	0	0	0	0	0	0	0	0
Total	**16**	**4**	**12**	**2**	**0**	**0**	**1**	**11**	**2**

Table 2: Descriptive Statistics: Raw Measures of Affiliation with the Catholic Church and the Protestant Church, 2002-2016

Panel A: Catholic Church

	Scandal-Stricken	Statistics	*C.Scands*	Total Population	Number of Members		Measures of Affiliation (*rC.Affil*, in %)			
					Total	% of Total Population	*rC.Attendance*	*rC.Exit*	*rC.Entry*	*rC.Baptism*
Catholic Dioceses	No	p25	0.00	2,235,000	745,000	34.00	11.01	0.33	0.01	0.67
		p50	0.00	3,415,000	1,485,000	41.12	12,53	0.48	0.01	0.72
		p75	0.00	4,900,000	1,938,510	55.29	14.44	0.64	0.02	0.78
		mean	0.00	3,343,671	1,344,392	43.24	12.99	0.51	0.02	0.73
	Yes	p25	1.00	2,143,000	1,075,000	34.54	9.90	0.43	0.01	0.65
		p50	1.00	3,778,000	1,609,000	40.41	11.14	0.57	0.01	0.69
		p75	2.00	5,000,000	1,915,000	53.33	13.30	0.81	0.02	0.77
		mean	1.67	3,659,196	1,456,865	44.15	12.14	0.64	0.01	0.70
	All	p25	0.00	2,186,000	767,000	34.00	10.61	0.34	0.01	0.67
		p50	0.00	3,505,000	1,506,000	41.09	12.32	0.50	0.01	0.72
		p75	0.00	4,900,000	1,937,391	54.49	14.14	0.68	0.02	0.78
		mean	0.39	3,487,323	1,370,767	43.45	12.79	0.54	0.01	0.72

Panel B: Protestant Church

	Scandal-Stricken	Statistics	*P.Scands*	Total Population	Number of Members		Measures of Affiliation (*rP.Affil*, in %)			
					Total	% of Total Population	*rP.Attendance*	*rP.Exit*	*rP.Entry*	*rP.Baptism*
Protestant Regional Churches	No	p25	0.00	4,762,530	1,239,774	22.71	2.50	0.51	0.12	0.77
		p50	0.00	6,141,143	2,335,722	32.66	2.75	0.61	0.13	0.83
		p75	0.00	8,028,639	2,662,789	38.74	3.47	0.76	0.17	0.88
		mean	0.00	6,723,899	2,007,933	32.53	3.06	0.66	0.14	0.83
	Yes	p25	1.00	6,217,829	2,323,155	20.33	2.51	0.56	0.13	0.82
		p50	1.00	6,280,717	2,456,140	22.62	2.62	0.75	0.16	0.84
		p75	2.00	12,500,000	2,629,670	36.99	3.66	0.88	0.17	0.87
		mean	1.43	8,942,413	2,370,968	29.26	2.98	0.77	0.16	0.83
	All	p25	0.00	5,020,012	1,252,395	22.40	2.50	0.52	0.12	0.77
		p50	0.00	6,175,752	2,346,879	32.58	2.75	0.61	0.14	0.83
		p75	0.00	8,056,673	2,662,789	38.34	3.49	0.80	0.17	0.88
		mean	0.09	6,874,307	2,030,769	32.31	3.05	0.67	0.14	0.83

Note: Measures of affiliation are prior to standardization by the respective sample mean.

Table 3: Effects of Catholic Church Scandals on Affiliation with the Catholic Church

	(1a)	**(1b)**	**(2a)**	**(2b)**	**(3a)**	**(3b)**	**(4a)**	**(4b)**
	C.Attendance		***C.Exit***		***C.Entry***		***C.Baptism***	
C.Scands	-1.809**	0.375	14.09***	4.087**	-3.850**	1.308	-1.663***	-0.893
l1.C.Scands	-2.428**	0.438	3.740**	0.448	-5.962***	1.287	-1.712***	-0.672**
Y2003		0.0599		8.122***		4.897		-2.473***
Y2004		-0.242		-6.664**		9.897**		-1.658
Y2005		-3.616		-15.37***		46.72***		-2.336
Y2006		-5.426**		-19.10***		39.91***		-5.686**
Y2007		-7.884**		-10.42**		38.61***		-5.317*
Y2008		-10.38**		9.020***		45.91**		-3.8
Y2009		-13.28***		12.99***		8.981		-7.422**
Y2010		-17.24***		46.93***		-3.452		-7.883*
Y2011		-19.34***		17.58***		-7.94		-8.306**
Y2012		-22.32***		12.76**		-15.43*		-10.68**
Y2013		-30.67***		59.40***		-11.14*		-11.83***
Y2014		-29.36***		97.82***		-16.90**		-10.98**
Y2015		-33.42***		66.27***		-24.21**		-10.12***
Y2016		-35.56***		50.05***		-22.58***		-7.243**
Year-FE	No	Yes	No	Yes	No	Yes	No	Yes
Diocese-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R2	0.02	0.77	0.09	0.83	0.02	0.35	0.07	0.29
N	405	405	405	405	405	405	405	405

Standard errors are clustered by dioceses. *p<0.10, **p<0.05, ***p<0.01.

Table 4: Effects of Institution-/ Individual-caused Catholic Church Scandals on Affiliation with the Catholic Church

	(1a)	**(1b)**	**(2a)**	**(2b)**	**(3a)**	**(3b)**	**(4a)**	**(4b)**
	C.Attendance		***C.Exit***		***C.Entry***		***E.Baptism***	
C.ins.Scands	-1.641**	0.262	12.81***	3.454**	-3.166**	1.64	-1.559***	-0.876
l1.C.ins.Scands	-2.422**	0.296	3.182*	0.273	-5.283***	1.764	-1.619***	-0.613*
C.idv.Scands	-4.980*	2.073*	36.08**	15.64**	-14.10**	-3.476	-3.253**	-1.017
l1.C.idv.Scands	-2.808	3.803**	16.75	6.741	-20.43**	-9.802	-3.727***	-1.938**
Y2003		0.0694		8.060***		4.858		-2.481**
Y2004		-0.258		-6.775**		9.941**		-1.657
Y2005		-3.637		-15.49***		46.78***		-2.333
Y2006		-5.509**		-19.64***		40.15***		-5.678**
Y2007		-8.030**		-10.77**		39.08***		-5.267*
Y2008		-10.38**		9.043***		45.90**		-3.801
Y2009		-13.32***		12.55**		9.077		-7.429**
Y2010		-17.34***		46.56***		-3.153		-7.858
Y2011		-19.48***		16.93**		-7.54		-8.280**
Y2012		-22.38***		12.64**		-15.24*		-10.66**
Y2013		-30.69***		59.04***		-11.10*		-11.84**
Y2014		-29.50***		97.27***		-16.45**		-10.94**
Y2015		-33.56***		65.66***		-23.76**		-10.09**
Y2016		-35.71***		49.44***		-22.09***		-7.203**
Year-FE	No	Yes	No	Yes	No	Yes	No	Yes
Diocese-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R2	0.03	0.77	0.1	0.83	0.02	0.35	0.08	0.29
N	405	405	405	405	405	405	405	405

Standard errors are clustered by dioceses. *p<0.10, **p<0.05, ***p<0.01.

Table 5: Effects of Catholic Scandals on the Protestant Church

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
	P.Attendance		*P.Exit*		*P.Entry*		*P.Baptism*	
C.Scands	-0.614**	-0.275	0.705	1.004	1.079**	1.364***	-0.480*	-0.0185
l1.C.Scands	-0.174	-0.201	0.154	-0.0753	-0.148	0.936**	-0.692***	-0.314*
P.Scands	0.104	0.0893	3.588	-0.946	1.273	1.73	-1.257*	0.422
l1.P.Scands	-1.188	-0.527	1.183	1.846	-2.859	0.13	0.145	1.021
Y2003		-0.701		2.94		4.611		-3.296***
Y2004		2.255		-17.29***		7.772**		-1.334
Y2005		-4.727**		-30.45***		20.67**		-2.137
Y2006		0.0208		-28.57***		13.75***		-5.137**
Y2007		2.101		-21.60***		13.10***		-8.406***
Y2008		-2.676**		0.987		3.032		-6.955**
Y2009		0.466		-9.063**		5.184**		-8.467**
Y2010		-4.153		-13.30**		5.184		-10.72**
Y2011		-1.394		-12.32**		-3.823		-8.976**
Y2012		-3.231		-10.61**		-2.294		-10.61***
Y2013		-9.281**		12.76**		-4.115		-11.83***
Y2014		-3.812		81.14***		-18.15***		-13.85***
Y2015		-10.02**		40.69***		-19.04***		-12.48***
Y2016		-7.342**		28.67***		-18.35***		-7.801**
Year-FE	No	Yes	No	Yes	No	Yes	No	Yes
Regional Church-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R2	0	0.21	0	0.87	0.01	0.52	0.04	0.42
N	285	285	285	285	285	285	285	285

Standard errors are clustered by regional churches. *p<0.10, **p<0.05, ***p<0.01.

Table 6: Effects of Institution-/ Individual-caused Catholic Church Scandals on Affiliation with the Protestant Church

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
	P.Attendance		***P.Exit***		***P.Entry***		***P.Baptism***	
C.ins.Scands	-0.373	-0.0921	-0.333	0.865	1.593*	1.569***	-0.51	-0.206
l1.C.ins.Scands	-0.148	-0.172	-0.369	-0.0431	0.366	1.259**	-0.785**	-0.419*
C.idv.Scands	-3.712	-2.437*	11.53	3.045	-2.68	1.107	-0.656	1.635
l1.C.idv.Scands	-0.558	-0.731	7.638	-0.419	-7.497	-3.638	0.64	1.254
P.Scands	-0.214	-0.132	4.6	-0.714	0.997	1.824	-1.296**	0.56
l1.P.Scands	-1.324	-0.658	2.227	1.893	-3.661	-0.294	0.262	1.227
Y2003		-0.67		2.903		4.57		-3.308***
Y2004		2.302		-17.33***		7.809**		-1.378
Y2005		-4.679**		-30.49***		20.73**		-2.186
Y2006		0.183		-28.71***		13.81***		-5.273**
Y2007		2.189		-21.63***		13.39***		-8.545***
Y2008		-2.679*		0.992		3.004		-6.956**
Y2009		0.602		-9.208**		5.743**		-8.548***
Y2010		-4.14		-13.37**		4.875		-10.65**
Y2011		-1.261		-12.45**		-3.855		-9.065**
Y2012		-3.233		-10.60**		-2.239		-10.62***
Y2013		-9.188**		12.65**		-4.231		-11.87***
Y2014		-3.673		81.02***		-18.05***		-13.98***
Y2015		-9.855**		40.57***		-18.86***		-12.64***
Y2016		-7.211**		28.58***		-18.17***		-7.944**
Year-FE	No	Yes	No	Yes	No	Yes	No	Yes
Regional Church-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R2	0	0.2	-0.01	0.87	0	0.52	0.02	0.42
N	285	285	285	285	285	285	285	285

Standard errors are clustered by regional churches. *p<0.10, **p<0.05, ***p<0.01.

Acknowledgments

We thank all participants at the 44th Annual Congress of the European Accounting Association for their helpful comments.

Data Source

German Bishops Conference 2001-2016 and the Evangelical Church in Germany 2001-2016.

Conclusion

The doctoral thesis at hand contains three empirical essays on the effects of tax policies on different economic agents (individuals, firms, and governments) in three distinct areas of taxation widely overlooked by prior empirical research. Specifically, the first essay included in this thesis studies firms' responses to threshold-dependent tax enforcement policies. The second essay studies intrastate tax competition between local governments and profit shifting by firms to domestic tax havens. Finally, the third essay examines the effects of scandals on organizational affiliation and competition in a setting where religious organizations levy church taxes on their members.