Beyond the Notch: Revenue Manipulation and Business Splitting under Simplified Tax System

Yulia Kuchumova[*], Alina Ozhegova[†], Kristina Odud[‡]

January 8, 2026

Abstract

Simplified tax regimes with explicit eligibility thresholds are a common tool which reduces tax burden but distorts firms' behavior, creating local bunching just below the threshold. However, we show that when firms have access to business splitting their responses extend far beyond the local margin. Our model demonstrates that splitting reduces local bunching but amplifies distortions elsewhere in the distribution. Exploiting a reform in Russia's Simplified Tax System that significantly raised the revenue threshold, allowing us to use the post-reform distribution as a counterfactual, we detect an excess mass far below the threshold, providing the evidence that extensive-margin responses matter.

Keywords: size-based taxation; business splitting; notch; tax evasion

JEL classification: H25, H32, H26, D21, D22

[*]Department of Economics, National Research University Higher School of Economics in St. Petersburg
[†]Department of Economics, Aalto University School of Business & Helsinki GSE
[‡]Business Development and Efficiency Department, Alfa Bank

1 Introduction

Introducing some tax reliefs for small firms to reduce their tax burden and accounting costs is a common practice for most governments.[1] To implement such policies, governments often rely on explicit eligibility thresholds—such as limits on revenue, employment, or assets—that separate firms into different tax regimes. By design, these size-dependent rules create discontinuities in firms' tax obligations, which in turn generate behavioral responses. The literature has documented that such policies lead to local bunching of firms just below eligibility cutoffs, meaning that firms underreport revenues or adjust their real activity (Best et al., 2015; Kleven and Waseem, 2013; Harju et al., 2019). In practice, firms may also change organizational structures to avoid crossing a threshold. Because these responses are harder to detect and measure, they are often overlooked, leaving the full extent of distortions from size-based regulation insufficiently understood.

In this paper, we study how firms respond to size-based tax regimes when they have access not only to standard forms of underreporting but also to extensive-margin strategies such as business splitting. Our goal is to measure how these additional responses alter the distribution of firms around eligibility thresholds and to assess the implications for both theory and empirical methods. Specifically, we examine whether ignoring such extensive-margin adjustments leads to an underestimation of the behavioral distortions induced by preferential tax regimes.

We develop a simple theoretical model that incorporates both intensive-margin, underreporting, and extensive-margin, business splitting. The model predicts that when splitting is feasible, fewer firms bunch locally at the threshold, while more firms shift far below it, creating a larger missing mass above the cutoff and excess mass across a broader range. Using a novel dataset and an extended bunching methodology (Kosonen and Matikka, 2023), we confirm these predictions: firms'

[1] Table 2 in Appendix A in Sharma et al. (2025) provides a non-exhaustive list of countries that, as of May 2024, apply a threshold-type of dual-regime tax system, which provide a preferential treatment to a small-sized firms.

responses to threshold-based regulation are substantially larger than standard local bunching estimates suggest, and adjustments occur not only at but also well below the threshold. The paper provides the first empirical evidence that accounting for business splitting is essential for accurately measuring the distortions created by threshold-based tax systems.

We study these questions in the context of Russia's *Simplified Tax System (STS)*, which has provided small businesses with preferential treatment since 2002 and currently covers nearly half of all firms. Eligibility depends on several criteria, most notably that a firm's annual revenue does not exceed a fixed threshold. By design, this creates a revenue notch: firms just above the threshold face a sharp increase in tax liability and compliance costs. In 2017, Russia sharply increased this revenue threshold, from 79.7 mln rubles to 150 mln rubles, providing an opportunity to observe how firms' distributions change when the notch is moved upward.

In Russia, to take advantage of the STS, in addition to under-reporting revenue, some companies and entrepreneurs use such illegal practice as business splitting. For instance, in 2023, several criminal cases were filed in Russia against prominent bloggers for evading taxes on a massive scale, with the unpaid taxes amounting to approximately 1,320 billion rubles.[2] The bloggers used business splitting to stay within established STS thresholds and pay taxes at significantly reduced rates. Thus, business splitting gives firms an additional opportunity to illegally qualify for the STS and therefore evade taxes.

Business splitting is not unique to the Russian setting. In Japan, the VAT policy threshold induces firms to restructure their organizations by splitting some of their member corporations (Onji, 2009). In the US, some firms might split after the Tax

[2]Blogger Sasha Mitroshina: who she is and how much she earns (URL: `https://www.kommersant.ru/doc/5887863`); "The marathon" with the authorities is over: Elena Blinovskaya was finally accused of tax evasion and legalization (URL: `https://www.kommersant.ru/doc/6592155`); Valeria (Lerchek) and Artem Chekalin: who they are and how much they earn (URL: `https://www.kommersant.ru/doc/5863388`)

Cuts and Jobs Act (TCJA) to benefit from the Qualified Business Income (QBI) deduction (Gale et al., 2019). Additionally, the adoption of the economic nexus thresholds by states (after 2018) creates incentives for businesses to stay small and hence to split business. In a similar vein, special tax regimes also affect the choice of organizational form (Elschner, 2013), which is a related margin because it means changing the firm structure. However, overall, empirical evidence of business splitting is limited because it is difficult to detect and measure.

Our empirical analysis relies on the Russian Financial Statements Database (Bondarkov et al., 2025), which covers nearly the entire population of incorporated firms. Using firm-level financial data from 2016 to 2019, we examine reported revenues around both the pre- and post-reform thresholds. To estimate behavioral responses, we adapt the extended bunching methodology of Kosonen and Matikka (2023), which compares pre- and post-reform revenue distributions over a broad range below the threshold. The large 2017 reform provides the counterfactual needed to identify excess mass attributable to the notch.

In our analysis, we compare firms eligible for the STS (treatment group) with those that are not (control group). For the treatment group, the revenue distribution reveals a broad excess mass of firms not only just below but but also well below the pre-reform threshold, relative to the post-reform years (2017–2019). By contrast, the revenue distribution of the control group remains stable before and after the reform. This pattern indicates that, in addition to local bunching, firms engage in business splitting. Extended bunching estimates are also consistently larger than those obtained using the standard local method, confirming that ignoring extensive-margin responses understates the true distortions induced by the regime.

Next, we examine the reaction of firms to the large shift in the threshold that occurred in 2017. A comparison of the post-reform distributions (2017, 2018, and 2019) around the new threshold with the pre-reform distribution shows that a new concentration below the updated threshold starts to appear in 2017 and becomes sizable and evident only in 2018, with a slight increase in 2019. That is, firms did not immediately adapt to such a large change in the threshold, and it took them almost a year to adjust. This indicates inertia in the responses of the firms. Our bunching coefficient estimates at the new threshold for 2017, 2018, and 2019 confirm this observation. Interestingly, at the new threshold, the bunching coefficient is quite large, reflecting that even though there is a lower density of firm and corresponding lower excess mass compared to those at the old threshold, the intensity of bunching at the new threshold is still significant.

The paper contributes to the literature on size-based taxation and firm behavior Kleven and Waseem (2013); Best et al. (2015); Almunia and Lopez-Rodriguez (2018); Harju et al. (2019); Bachas and Soto (2021); Zanoni et al. (2025). There is a large body of work that uses local bunching methods to study behavioral responses to size-based regulations. The pioneering study by Kleven and Waseem (2013) shows that self-employed taxpayers in Pakistan actively respond to personal income tax notches. Best et al. (2015) document sharp bunching of Pakistani firms at the kink separating turnover and profit tax regimes. Other studies extend this evidence across contexts, showing bunching at the VAT exemption threshold in Finland (Harju et al., 2019), at corporate profit thresholds in Costa Rica (Bachas and Soto, 2021), at tax regime switches in Ecuador (Zanoni et al., 2025), and at income thresholds in Italy that also generate negative spillovers for competitors (Di Marzio et al., 2025). We build on this literature by showing that focusing solely on local bunching underestimates the true extent of firms' responses. Our analysis demonstrates that when extensive-margin strategies such as business splitting are available, firms shift not only just below the threshold but also far below it.

The paper most closely related to ours is Onji (2009), who shows that the introduction of a VAT threshold in Japan induced firms to restructure through splitting. We build on this insight by developing a theoretical model that incorporates business splitting as an extensive-margin response and by showing how it alters the distribution of firms around thresholds. Our work extends Onji (2009) in several important ways. First, whereas the Japanese study relies on survey data from publicly listed companies with information on their subsidiaries, we use population-wide administrative data on all Russian firms, including those without observable affiliations. Second, while Onji applies kernel density estimation and focuses narrowly on the size distribution just below the VAT threshold, we employ an extended bunching approach to analyze the entire revenue distribution below the cutoff. Third, we link the empirical analysis to a model that highlights how the availability of business splitting changes firms' behavioral responses and reduces the amount of local bunching observed.

The paper also complements the theoretical literature on the optimal design of dual-regime tax systems (Keen and Mintz, 2004; Dharmapala et al., 2011; Wei and Wen, 2019, 2023; Sharma et al., 2025). Previous literature typically emphasizes intensive-margin adjustments, such as underreporting or reduced production, and derives optimal thresholds by weighing these distortions against administrative savings. Our contribution is to document that firms also respond at the extensive margin by splitting, which amplifies distortions and reshapes the distribution of firms far below the threshold. Ignoring this additional channel means that theoretical models may underestimate efficiency costs and, as a result, recommend thresholds and tax parameters that are not truly optimal.

This paper proceeds as follows. Section 2 provides the relevant institutional details. Section 3 presents the theoretical framework. Section 4 presents data and descriptive statistics. Section 5 describes our empirical methods. Section 6 presents the results and discusses the findings. Section 7 concludes.

2 Institutional details

2.1 Simplified tax system: characteristics and conditions to use

In Russia, there are several tax regimes for organizations and entrepreneurs: a general one and several specialized ones. By default, all firms (and entrepreneurs) have to pay taxes according to the general tax regime, which is the most comprehensive and complicated in terms of calculating taxes and the record keeping process. Under the general schedule, firms are liable for profit tax, VAT, and property tax. During the period from 2011 to 2019, the basic profit tax rate was 20%, the VAT rates were 0%, 10%, and 18%, and the maximum property tax rate for firms was 2.2% [3]

Additionally, if a firm satisfies certain criteria then it can choose a specialized tax regime (including a simplified tax system). The switch from the general regime to the specialized regime is voluntary. To make the switch, a firm needs to submit an application.

The Simplified Tax System in Russia was introduced in 2002 by Federal Law 104-FZ (dated 24.07.2002). This regime was developed specifically to support small and medium-sized businesses, as it allows reducing the tax burden under certain restrictions. Unlike the General Tax System (GTS) under the simplified tax system (STS), taxpayers are released from paying value added tax (VAT), have the right to carry out "simplified accounting", i.e. to reflect revenue and expenses on a cash basis, and are obliged to file a tax declaration only once a year, after the end of the reporting period. Individual entrepreneurs and organizations are also relieved from paying tax on property used in entrepreneurial activities, except for property taxed on the basis of its cadastral value.

Additionally, when a firm switches to the STS, it can choose which tax base

[3] The basic profit tax rate was 20% in the period 2009-2024. The VAT rates were 0%, 10% and 18% in the period 2004-2018. The property tax rate for firms is set by the laws of regions with a maximum rate of 2.2%.

it will use. The tax base can be equal to either revenue or revenue reduced by the amount of expenses. The tax rate then depends on the tax base chosen by the firm. Under the 'revenue' tax base, the tax rate is 6%. Under the 'revenue minus costs' tax base, the tax rate is 15%. Note also that taxpayers have the right to independently choose the tax base, except for certain cases (Article 346.14 of the Tax Code of the Russian Federation). For example, if a business is organized under a simple partnership agreement or trust management agreement, only ''revenue minus costs'' may be selected as the tax base.

An important advantage of the STS is that the firm's tax burden is reduced. Instead of a tax rate of 20% of the amount of taxable income (i.e., revenue minus costs) applied under the GTS, under the STS the taxpayer is obliged to pay tax of 6% of the total amount of revenue, or 15% of the amount of taxable income (i.e., revenue minus costs). Moreover, in some Russian regions the tax rate is reduced to 0% in the first year after the registration of an individual entrepreneur or LLC and the application of the STS.

There are, however, a couple of downsides of using the STS for a firm. First, when a firm uses the STS, many contractors would not be able to get VAT refund, which reduces their motivation to deal with such a firm. Second, even if a firm incurs losses in a given tax period (when it uses "revenue minus costs" tax base) it still has to pay a minimal tax which is equal to 1% of firm's revenue .

To apply for the STS, a company or individual entrepreneur must meet the criteria established by law, which are:

1. The total annual revenue should be less or equal than a certain threshold (details on threshold values are presented in Table 1).

2. The average annual number of employees should be less or equal a certain threshold (details on threshold values are presented in Table 1).

3. The residual value of fixed assets in the reporting period should be less or

equal a certain threshold (details on threshold values are in Table 1).

4. The share of the equity capital of the organization owned by third legal entities must not exceed 25%.

5. The type of economic activity carried out by the company does not fall under prohibited category (a complete list of economic activities, for which the use of the STS is prohibited, is presented in Appendix A).

Table 1: Threshold defined by the simplified tax system before and after the 2017 tax reform

Period	Total annual revenue threshold (thousand RUB)	Average annual number of employees threshold	Residual value of fixed assets threshold (thousand RUB)
2016	79,740	100	100,000
2017–2018	150,000	100	150,000

Source: Article 346.12 of the Tax Code

By its design, the STS criterion on revenue creates the discontinuity (notch) in the tax burden: companies exceeding the established revenue threshold are obliged to switch to the general tax system and pay 20% of revenue instead of the established reduced rate under the STS. That is, in the case of exceeding revenue threshold, the increase in the average revenue tax rate is discontinuous and sizable. It is 5% (from 15% in the case of the 'revenue minus costs' tax base to 20%) and higher (if regional reductions take place) for firms with the 'revenue minus cost' tax base. In the case of the 'revenue' tax base, the notch can be even larger: given that the 'revenue' tax base is a voluntary choice, it should be more favorable for the firm. Moreover, if the revenue threshold is exceeded, in addition to the increase in the amount of taxes, additional obligations arise: bookkeeping becomes more complex (which entails additional costs for its maintenance), the number and frequency of tax returns and advance tax payments increases, and the need to pay value added tax arises. Therefore, in general, the whole process becomes more complex and expensive. Thus, a notch at the revenue threshold arises not only because of the jump in the tax liabilities, but also because of the increase in general administrative cost. Note that the same logic is applicable to the other thresholds (for the number of employers and for the residual value of fixed assets). However, in this paper, we mainly focus on the notch associated with the revenue threshold.

It is important to note that in 2017 the regulations on the STS were partially modified. Prior to 2016, the revenue threshold was indexed annually: starting from 2014, the amount of 60,000 thousand rubles was increased annually in accordance with the established inflation coefficients. Correspondingly, in 2016, the threshold amounted to 79,400 thousand rubles. However, in 2017, the revenue threshold for the STS was significantly increased to 150,000 thousand rubles, as shown in Figure 1. Firms with an annual revenue of up to 150,000 thousand rubles were eligible to use the STS, which represented an 89% increase. It should be noted that starting from 2017 indexation of the revenue threshold had been suspended until 2021. Additionally, in 2017, the threshold for the residual value of fixed assets was significantly increased by 50% up to 150,000 thousand rubles. This sizable change enabled a greater number of companies to apply for the STS.

2.2 Illegal practices of using the STS in Russia

In an effort to retain the right to use the simplified tax system, some companies try to decrease their revenues by various means. One typical method is to under-report sales or over-report costs. Another method used in practice is to operate a business through separate legal entities or individual entrepreneurs, effectively splitting the business. Although, splitting a business is generally permissible, to be legal it should be conducted in compliance with all applicable laws and reg-

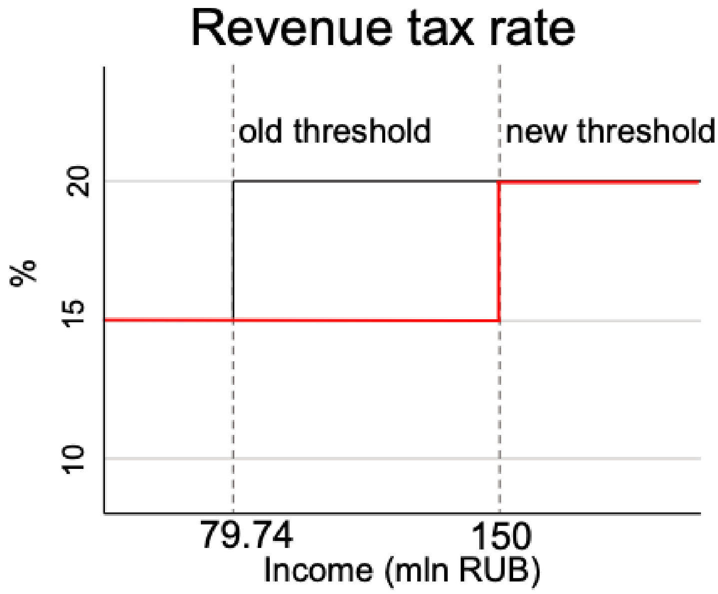


Figure 1: Revenue Tax Rates for Different Levels of Revenue Before and After the Tax Reform In 2017

ulations. According to Clause 1 of Article 54.1 of the Tax Code of the Russian Federation, if the act of splitting is carried out with the intention of unjustified tax benefits due to the application of special regimes, privileges, and tax savings, it is a direct violation of the law. It is worth noting that the term 'splitting' is not defined or applied within the framework of the legislation. Formally, this is done by registering several interrelated enterprises in one's own name or in the name of relatives and friends.

The Russian Federal Tax Service acknowledges that there are instances of companies unlawfully retaining the right to use the simplified tax system. To address this issue, the Federal Tax Service conducts regular audits, as outlined in the letter from the Federal Tax Service.[4] In 2023, the Federal Tax Service inspections were primarily focused on companies that used the STS, especially those whose revenues approached the threshold values. Following inspections, several promi-

[4]Letter from the Federal Tax Service is available at: https://www.consultant.ru/document/cons_doc_LAW_317165/

nent bloggers have been charged with tax evasion on a massive scale. It has been found that the total amount of unpaid taxes in these cases was approximately 1,320 billion rubles[5].

These bloggers were found to have used a business splitting scheme that is not in compliance with established regulations, resulting in a significant reduction in tax payments. The scheme involves the creation of new companies under the STS by the spouse and other relatives of the blogger when the revenue of the original company (operating under the STS) approaches the threshold value. The revenue generated from advertising and other sources is then transferred to the accounts of these newly created companies. However, if revenue of the original company continued to be accounted for as before (without splitting business), the company would exceed the STS revenue threshold and would be obliged to apply the general tax system. This could result in a significant increase in the total amount of taxes payable: firstly, income tax would have to be paid at a higher rate (15% for sole proprietorship and 20% for LLC) instead of the reduced rate under the STS, and secondly, VAT would also have to be paid. Therefore, this behavior of the bloggers falls under the signs of illegal splitting of business.

3 Model

To examine the effects caused by the business splitting opportunity available to firms, we develop a theoretical framework characterizing firms' behavior induced by the STS notch. Our model builds on Carvalho (2024) and Dharmapala et al. (2011) but with some simplifications. Given that our goal is to examine the incentives created by the STS threshold without delving into welfare analysis, we retain only the main assumptions of the baseline models and omit preferences for

[5]Blogger Sasha Mitroshina: who she is and how much she earns (URL: https://www.kommersant.ru/doc/5887863); "The marathon" with the authorities is over: Elena Blinovskaya was finally accused of tax evasion and legalization (URL: https://www.kommersant.ru/doc/6592155); Valeria

the choice of the production sector and dynamic characteristics concerning firm entry and exit.[6] In what follows, we first consider tax evasion modeled in the traditional way, corresponding to intensive margin responses. Then we model a situation in which firms have the additional opportunity to evade using business splitting, representing extensive margin responses. We characterize how business splitting opportunity modifies firms' behavior around the tax notch.

Consider a large number of firms that produce a single homogeneous good and sell their products to consumers. Assume that the demand for the good is perfectly elastic and the producer price of the good is normalized to one. Firms are heterogeneous and characterized by their productivity level a. A firm of type a produces y units of final good at cost $\varphi(y,a)$. Hence, the before-tax profit of a firm is $y-\varphi(y,a)$. The cost is strictly increasing and strictly convex in revenue y, i.e., $\varphi_y^{'}(y,a)$ and $\varphi_{yy}^{''}(y,a)$ are positive, but high-productivity firms incur a lower total and marginal cost of generating revenue, such that $\varphi_a^{'}(y,a)$ and $\varphi_{ya}^{''}(y,a)$ are negative. Additionally, assume that as a firm grows, the effect of an increase in a on the marginal cost becomes larger relative to the effect of an increase in y, specifically, assume that $\frac{-\varphi_{ya}^{''}(y,a)}{\varphi_{yy}^{''}(y,a)}$ is a non-decreasing function of y.[7]

Firms have to pay taxes. There are two different tax regimes. Under the general tax regime, a firm remits tax on profit at the rate t_g and also incurs compliance costs $\theta_g > 0$, associated with an accounting complexity. Under the simple tax regime (the so-called Simplified Tax System, i.e., STS), a firm remits a tax on profit at the rate $t_s < t_g$ and the compliance costs are small and normalized to zero. However, the simple tax regime can only be applied by a firm, whose revenue is below a given threshold, $\overline{Y}$. Although in practice, under the STS in Russia, a

[6] Although Carvalho (2024) talks about entrepreneurs, the same framework is also suitable for describing firm's behavior because a model with the choice of labor workers, such as Alvarez, Pessoa, Souza (2022), can be transformed into the model of Carvalho (2024). They are mathematically equivalent, as one can be converted into the other by replacing the optimization variable.

[7] Note that this assumption is valid for iso-elastic function, $\varphi(y,a) = \frac{a}{1+\frac{1}{e}}\left(\frac{y}{a}\right)^{1+\frac{1}{e}}$, which is traditionally used in the bunching literature (Harju, Matikka, Rauhanen, 2019).

firm may choose between the 'profit' and 'revenue' tax base, in a model we only consider the 'profit' tax base case, since this simplifies the analysis but does not significantly change the conclusions.

3.1 Traditional tax evasion case

However, firms may evade paying some taxes. Evading taxes is costly, and the evasion cost, $c(e)$, is a strictly increasing and strictly convex function of evasion, e, i.e., $c^{'}(e)$ and $c^{''}(e)$ are positive. For now, we will consider only this option of evading taxes, which creates continuous and local incentives for evasion. Later, we will also add an option of evasion by means of business splitting, which creates discontinuous evasion response.

Thus, the firm's after-tax profit under the tax regime $k = \{g, s\}$ for a given revenue, y, and evasion, e, is

$$\Pi_k(y, e|a) = y - \varphi(y, a) - T_k(y, e) - c(e), \tag{1}$$

where $T_g(y, e|a) = t_g[y - \varphi(y, a) - e] - \theta_g$ and $T_s(y, e|a) = t_s[y - \varphi(y, a) - e]$. A firm can apply the simple tax regime if its reported revenue is below the threshold, i.e. $y - e < \overline{Y}$.

Given that $T_s(y, e|a) < T_g(y, e|a)$ for any (y, e, a), it follows that $\Pi_s(y, e|a) < \Pi_k(y, e|a)$ for any (y, e, a). That is, a firm will always choose the simple tax regime over the general tax regime whenever possible. Based on this inference, the firm's optimization problem can be formulated as:

$$\underset{(y,e)}{max}\, \Pi(y, e|a) = \begin{cases} y - \varphi(y, a) - T_g(y, e) - c(e), & \text{if } y - e > \overline{Y}, \\ y - \varphi(y, a) - T_s(y, e) - c(e), & \text{if } y - e \leq \overline{Y}. \end{cases} \tag{2}$$

The simple tax regime imposes a lower tax burden on a firm, and hence its existence creates a notch in tax obligations at the revenue threshold $\overline{Y}$, providing incentives to reduce reported revenue below the threshold $\overline{Y}$. To understand exactly how firms respond to these incentives, we need to define the solution to the firm's problem.

To proceed, let $(y_g^*(a), e_g^*)$ denote the optimal revenue and evasion if only general tax regime existed, that is, the unconstrained solution of $\underset{(y,e)}{max}[y - \varphi(y,a) - T_g(y,e) - c(e)]$, which is characterized by the following FOCs:

$$\begin{cases} \varphi_y^{'}(y_g^*, a) = 1, \\ c^{'}(e_g^*) = t_g. \end{cases} \tag{3}$$

Similarly, let $(y_s^*(a), e_s^*)$ denote the optimal revenue and evasion if only simple tax regime existed, that is, the unconstrained solution of $\underset{(y,e)}{max}[y - \varphi(y,a) - T_s(y,e) - c(e)]$, which is characterized by the following FOCs:

$$\begin{cases} \varphi_y^{'}(y_s^*, a) = 1, \\ c^{'}(e_s^*) = t_s. \end{cases} \tag{4}$$

As we can see from the two FOC systems, for a given productivity a, the optimal revenue levels are the same under both tax regimes, $y_s^*(a) = y_g^*(a)$. Additionally, they are socially efficient, as $\varphi_y^{'}(y_s^*, a) = 1$ and $\varphi_y^{'}(y_g^*, a) = 1$, and increase in productivity a. Furthermore, since the tax evasion depends only on the tax rate, the simple tax regime by itself creates fewer incentives for evasion than the general tax regime by itself, $e_s^* < e_g^*$. However, as we will see later, the existence of the tax notch at the revenue threshold, $\overline{Y}$, creates additional and substantial incentives for tax evasion.

The discussed results are proved in the following lemma.

Lemma 1. *a) The optimal unconstrained revenues under both general and simple tax regimes, defined by the systems of equations (3) and (4),* *are socially efficient.*

b) The tax evasion depends only on the tax rate. The tax evasion under the simple tax regime by itself is lower than the tax evasion under the general tax regime by itself, $e_s^* < e_g^*$.

c) The optimal unconstrained revenue is an increasing function of productivity a, $\frac{\partial y_s^*(a)}{\partial a} = \frac{\partial y_g^*(a)}{\partial a} = -\frac{\varphi''_{ya}(y_s^*,a)}{\varphi''_{yy}(y_s^*,a)} > 0$.

Proof. See proof in Appendix B.

According to Lemma 1, $y_s^*(a)$ is increasing in the productivity. Hence, as productivity a increases, $y_s^*(a) - e_s^*$ can become greater than the revenue threshold $\overline{Y}$. Let a^* denote the productivity level such that

$$y_s^*(a^*) - e_s^* = \overline{Y}. \tag{5}$$

But, at the threshold $\overline{Y}$, the firm faces the tax notch, since it has to switch to the general tax regime, which creates additional incentives to reduce the reported revenue. Given these incentives, firms with productivity $a > a^*$ may find it optimal to report $\overline{Y}$ instead of $y_s^*(a) - e_s^*$. To explore this case, let $(\hat{y}_s(a), \hat{e}_s(a))$ denote the solution of $\underset{(y,e)}{max}[y - \varphi(y,a) - T_s(y,e) - c(e)]$ under constraint $y - e = \overline{Y}$. The Lagrangian for this problem is $L = y - \varphi(y,a) - T_s(y,e) - c(e) + \lambda(\overline{Y} - y + e)$ and the FOCs are

$$\begin{cases} \varphi'_y(\hat{y}_s, a) = 1 - \frac{\lambda}{1-t_s}, \\ c'(\hat{e}_s) = t_s + \lambda, \\ \hat{y}_s - \hat{e}_s = \overline{Y}, \\ \lambda > 0. \end{cases} \tag{6}$$

Several inferences follow from these FOCs. The first FOC in the above system implies that $\hat{y}_s(a)$ is not socially efficient and it is lower than $y_s^*(a)$. The second FOC implies that $\hat{e}_s(a)$ is greater than e_s^*. The third FOC tells us that the firm's reported revenue is exactly equal to the threshold level, $\hat{y}_s - \hat{e}_s = \overline{Y}$, that is, such a firm bunches at the threshold. Thus, the firm that bunches at the threshold, uses two channels to reduce its reported revenue: first, the firm reduces its production to an inefficient level, and second, it increases its evasion. That is, the notch produces both real and reporting effects, which are both distortionary. Moreover, we can show that $y_s^*(a) - \hat{y}_s(a)$ and $\hat{e}_s(a) - e_s^*$ increase with a. Hence, the higher the productivity a, the more significant these effects and the distortions caused by them.

The following lemma summarizes these results.

Lemma 2. *a) The revenue of a bunching firm, defined by the system of equations (6), is lower than the efficient level, $\hat{y}_s(a) < y_s^*(a)$.*

b) The evasion of a bunching firm, $\hat{e}_s(a)$, is greater than e_s^*.

c) *The revenue and the tax evasion of a bunching firm are increasing functions of productivity a,* $\frac{\partial \hat{y}_s(a)}{\partial a} = \frac{\partial \hat{e}_s(a)}{\partial a} = -\frac{\varphi''_{ya}(\hat{y}_s,a)}{\varphi''_{yy}(\hat{y}_s,a)+\frac{1}{1-t_s}c''(\hat{y}_s(a)-\overline{Y})} > 0$.

d) The distortions, caused by the notch, increase with productivity level, that is, $y_s^*(a) - \hat{y}_s(a)$ and $\hat{e}_s(a) - e_s^*$ increase with a.

Proof. See proof in Appendix B.

Let us now define the maximal profit of a firm depending on its productivity.

The maximum profit of a firm under the simple tax regime with productivity $a \le a^*$ is

$$\Pi_s^*(a) = \Pi_s(y_s^*(a), e_s^*) = (1-t_s)(y_s^* - \varphi(y_s^*, a)) + t_s e_s^* - c(e_s^*).$$

The maximum profit of a bunching firm with productivity $a > a^*$ is

$$\hat{\Pi}_s(a) = \Pi_s(\hat{y}_s(a), \hat{e}_s(a)) = (1-t_s)(\hat{y}_s - \varphi(\hat{y}_s, a)) + t_s\hat{e}_s - c(\hat{e}_s).$$

The maximum profit of a firm under the general tax regime with productivity a is

$$\Pi_g^*(a) = \Pi_g(y_g^*(a), e_g^*) = (1-t_g)(y_g^* - \varphi(y_g^*, a)) + t_g e_g^* - c(e_g^*) - \theta_g.$$

Note that for $a = a^*$, $\hat{\Pi}_s(a^*) = \Pi_s^*(a^*)$ because at a^* by definition $y_s^*(a^*) - e_s^* = \overline{Y}$. Additionally, as we discussed, $\Pi_g^*(a) < \Pi_s^*(a)$. Therefore, for a firm with productivity a^*, $\Pi_g^*(a) < \Pi_s^*(a) = \hat{\Pi}_s(a^*)$, that is, bunching is preferable to switching to the general tax regime.

However, the higher the productivity a $(a > a^*)$ of a bunching firm, the greater the distortions from inefficient production and increased evasion, and hence the greater the reduction in the firm's profit relative to the efficient production level. Indeed, if we apply the second-order Taylor approximation, we can obtain $\hat{\Pi}_s(a) \approx \Pi_s^*(a) - (1-t_s)\varphi_{yy}^{''}(y_s^*, a)\frac{(y_s^* - \hat{y}_s)^2}{2} - c^{''}(e_s^*)\frac{(\hat{e}_s - e_s^*)^2}{2}$, which represents that the reduction in the profit of a bunching firm increases with the distortions, i.e., $y_s^* - \hat{y}_s$ and $\hat{e}_s - e_s^*$. As a result of this, as productivity a gets higher, the profit of a bunching firm may become equal to the profit that can be earned under the general tax regime with optimal production and evasion. Consequently, the firms with such level of productivity and greater will choose the general tax regime. Let $\hat{a}$ define this productivity, which is characterized by the following equation:

$$\hat{\Pi}_s(\hat{a}) = \Pi_g^*(\hat{a}). \tag{7}$$

The following proposition proves that such a productivity level, $\hat{a}$, exists.

Proposition 1. *There exists productivity level $\hat{a}$ greater than a^* ($\hat{a} > a^* > 0$), defined by equation (7) such that i) a firm with productivity $a \geq \hat{a}$ chooses the general tax regime; ii) a firm with productivity $a^* < a < \hat{a}$ bunches at the threshold $\overline{Y}$; iii) firms with productivity $a < a^*$ choose the simple tax regime.*

Proof. See proof in Appendix B.

Proposition 1 completes the description of the solution to firm's problem and allows us to characterize the equilibrium sorting of firms into tax regimes in our economy. There will three cases: 1) firms with productivity $0 < a \leq a^*$ apply simple tax regime as their optimal reported income lower than the threshold; 2) firms with productivity $a^* < a < \hat{a}$ bunch at the threshold in order to use the simple tax regime, 3) firms $a \geq \hat{a}$ apply general tax regime. Also, it is important to remember that bunching not only reduces the revenue that can be collected by the tax authority but it also causes both real and reporting distortions.

3.2 Tax evasion through business splitting

Assume now that in addition to the traditional tax evasion at cost $c(e)$, the firm can also use business splitting. This allows the firm, at fixed cost b, to report its revenue net of tax evasion, $y-e$, as if it were from two separate firms, $y-e = x^{(1)} + x^{(2)}$.[8] A firm has incentives to use business splitting only if it helps it use the STS, that is, if $a > a^*$ and $x^{(1)} \leq \overline{Y}$ and $x^{(2)} \leq \overline{Y}$. This means that by using business splitting, a firm can apply the simple tax regime if $y-e \leq 2\overline{Y}$, which involves a fixed cost b. Note that, for simplicity of analysis, we will only allow the option to split business on two entities. Although in practice it might be possible to split business on more than two entities, there is a lack of understanding of how often this happens.

Then, if a firm with $a > a^*$ uses business splitting, its profit maximization problem becomes

$$\begin{gathered} \max_{(y,e)}[(1-t_s)(y-\varphi(y,a))-t_s e-c(e)-b] \\ s.t.\ y-e \leq 2\overline{Y}. \end{gathered} \tag{8}$$

If the constraint $y-e \leq 2\overline{Y}$ is not binding, i.e., $y-e < 2\overline{Y}$, then the solution to

[8]By cost b being fixed we mean that from a firm's point of view this cost does not depend on evasion and revenue. But, the fixed cost might be globally conditional on firm size: it could be easier to split for a relatively small firm than for a large firm. Here, for simplicity, we abstract from this aspect.

this problem is equal to $(y_s^*(a^*), e_s^*)$. That is, the firm's production is at efficient level. Note that for a not much larger than a^* the constraint is indeed not binding because at a^*, $y_s^*(a^*) - e_s^* = \overline{Y} < 2\overline{Y}$. Let a_b^* denote the value of productivity such that $y_s^*(a_b^*) - e_s^* = 2\overline{Y}$. Hence, for $a^* \leq a \leq a_b^*$ the solution of problem (8) is characterized by this case. Denote the maximal profit in this case by $\Pi_{sb}^*(a)$, which is equal to

$$\Pi_{sb}^*(a) = (1-t_s)(y_s^* - \varphi(y_s^*, a)) - t_s e_s^* - c(e_s^*) - b.$$

If $a > a_b^*$ and hence the constraint $y - e \leq 2\overline{Y}$ is binding, i.e., $y - e = 2\overline{Y}$, then let $(\hat{y}_{sb}(a), \hat{e}_{sb}(a))$ denote the optimal solution to problem (8). The FOCs characterizing $(\hat{y}_{sb}(a), \hat{e}_{sb}(a))$ are the same as system (6) except that the third FOC is now $\hat{y}_{sb} - \hat{e}_{sb} = 2\overline{Y}$. Because of this, $\hat{y}_{sb}(a) > \hat{y}_s(a)$ for any $a > a_b^*$.

Denote the maximal profit in this case by $\hat{\Pi}_{sb}(a)$, which is equal to

$$\hat{\Pi}_{sb}(a) = (1-t_s)(\hat{y}_{sb} - \varphi(\hat{y}_{sb}, a)) - t_s \hat{e}_{sb} - c(\hat{e}_{sb}) - b.$$

Furthermore, let $\Pi_{sb}^{max}(a)$ denote the maximal profit when business splitting is used, which is

$$\Pi_{sb}^{max}(a) = \begin{cases} \Pi_{sb}^*(a), & \text{if } a^* \leq a \leq a_b^*, \\ \hat{\Pi}_{sb}(a), & \text{if } a > a_b^*. \end{cases}$$

Given that business splitting involves paying fixed costs regardless of the amount of revenue being divided, its use may not always be optimal. In order to understand when it becomes optimal for a firm to use business splitting, let us compare the value of $\Pi_{sb}^{max}(a)$ with $\hat{\Pi}_s(a)$ for $a > a^*$. Although at a^*, $\Pi_{sb}^{max}(a) = \Pi_{sb}^*(a^*) < \Pi_s^*(a^*) = \hat{\Pi}_s(a^*)$ because $\Pi_{sb}^*(a^*) = \Pi_s^*(a^*) - b$, as productivity a increases, profit $\Pi_{sb}^{max}(a)$ grows faster than $\hat{\Pi}_s(a)$. Hence, there exists a productivity, call it $\underline{a_b}$, such that

$$\Pi_{sb}^{max}(\underline{a_b}) = \hat{\Pi}_s(\underline{a_b}).$$

In case if $\underline{a_b} \leq \hat{a}$, $\Pi_{sb}^{max}(\underline{a_b})$ is also greater than the profit under the general tax system $\Pi_g^*(\underline{a_b})$ because, as we have shown, for $a < \hat{a}$ bunching is preferable over the use of the general tax regime, i.e., $\hat{\Pi}_s(a) > \Pi_g^*(a)$. To ensure that $\underline{a_b} \leq \hat{a}$, we will now assume that fixed cost b is not too large.

Assumption 1. *Assume that fixed cost b is not too large so that* $\underline{a_b} \leq \hat{a}$.

Given assumption 1, for firms with productivity greater than $a > \underline{a_b}$, it is optimal to use traditional evasion in combination with business splitting rather than only traditional evasion. This means that the number of firms that bunch will decrease compared to the case when only traditional evasion was available. This is because only firms with productivity $a^* < a \leq \underline{a_b}$, where $\underline{a_b} < \hat{a}$, bunch at the threshold $\overline{Y}$, while firms with $\underline{a_b} \leq a < \hat{a}$, although eligible for the STS through business splitting, are no longer bunching, since their reported revenues $x^{(1)} \leq \overline{Y}$ and $x^{(2)} \leq \overline{Y}$ are relatively arbitrary.

Given that $\underline{a_b} < \hat{a}$, we have $\Pi_{sb}^{max}(\hat{a}) > \hat{\Pi}_s(\hat{a}) = \Pi_g^*(\hat{a})$ for $\underline{a_b} \leq a < \hat{a}$. This implies that firms with productivity $\hat{a}$ and greater would prefer to qualify for the simple tax regime through the use of business splitting and traditional evasion rather than apply the general tax regime. This means that the opportunity to use business splitting extends the ability to illegally apply the STS. However, at some point, even with business splitting, both real and reporting distortions, caused by the constrained optimization, arise. They are similar to what we have discussed in Lemma 2. As a increases, these distortions grow, and at some point the maximal profit with business splitting will become equal and then lower than the maximal profit from the use of the general tax regime. Let $\overline{a_b}$ denote the productivity level such that

$$\Pi_{sb}^{max}(\overline{a_b}) = \Pi_g^*(\overline{a_b}).$$

Note that productivity $\overline{a_b}$ is greater than $\hat{a}$, and it is this result that means that the opportunity to use business splitting extends the ability to illegally apply the STS.

The following proposition proves the results discussed in detail.

Proposition 2. *Given Assumption 1, there exist two productivity levels* $\underline{a_b}$ *and* $\overline{a_b}$ *such that* $\underline{a_b} < \hat{a}$ *and* $\overline{a_b} > \hat{a}$ *and that*

i) firms with productivity $a \in [a^*, \underline{a_b})$*, in order to apply the STS, use only traditional evasion and hence bunch at the threshold;*

ii) firms with productivity $a \in [\underline{a_b}, \overline{a_b})$*, in order to apply the STS, use business splitting in additional to traditional evasion and do not bunch;*

iii) firms with productivity $a \geq \overline{a_b}$*, apply the general tax regime.*

Proof. See proof in Appendix B.

Thus, the business splitting increases the illegal use of the STS while it decreases the number of firms bunching at the threshold $\overline{Y}$. Therefore, the standard (local) bunching method would under-estimate the amount of firms that use the STS through various extends of evasion and avoidance including business splitting.

4 Data and descriptive statistics

Our empirical analysis is based on the Russian Financial Statements Database (RFSD), an open-access harmonized dataset containing annual unconsolidated financial statements for nearly the entire population of Russian firms (Bondarkov et al., 2025). This database was only created in 2025. Unlike widely used commercial databases such as Orbis or Ruslana, the RFSD provides significantly broader coverage, including both filing and non-filing firms, enabling a more representative analysis of the firm population and more accurate estimations of behavioral responses to tax thresholds.

The RFSD integrates official administrative data from different sources. Firm registry details are sourced from the Unified State Register of Legal Entities (EGRUL) maintained by the Federal Tax Service (FNS), while firm-level financial statements are gathered from Rosstat (2011–2018) and the FNS (2019–2023).

A key advantage of the RFSD is that it also includes firms that failed to file financial statements despite being legally required to do so. Using the structure of Russian accounting rules, which requires prior-year figures to be included in current filings, the authors of the RFSD impute missing statements from forward-looking information. This imputation recovers an additional 5% of firm-year statements.

The database includes standard accounting reports: balance sheets, profit and loss statements, cash flow statements, and statements of equity. Each statement is subjected to articulation checks and internal consistency validation. Financial variables are harmonized across reporting forms and accounting standards, including changes in industry and legal form classifications over time. Firms are geocoded based on their incorporation address, and metadata flags indicate whether a firm is government-owned, strategic, under sanctions, or using simplified reporting forms.

For the purpose of the analysis, we focus on the years 2016–2019, which span the period immediately before and after the 2017 reform of the Simplified Tax System. In particular, we use firm-level data on total revenue, residual value of fixed costs, an industry code in terms of the Russian national classifier of economic activities OKVED and an indicator for whether the firm used the Simplified Accounting Statement (SAS).

Table 2 presents the number of active firms in Russia by the industry code from 2016 to 2019. The largest sector throughout the period is Wholesale and Retail Trade, with over half a million firms each year. Other prominent sectors include Construction, Manufacturing, and Professional, Scientific and Technical Activities, all of which saw steady growth across the years. In contrast, sectors such as Financial and Insurance Activities, Electricity and Gas Supply, and

Mining and Quarrying remained relatively small. The number of firms in Public Administration, Household Services, and Extraterritorial Organizations was minimal throughout. Overall, most sectors experienced moderate growth, particularly between 2018 and 2019, suggesting broad-based expansion in business activity during this period.

Additionally, Table 3 reports summary statistics for firms from 2016 to 2019. Average revenue peaked in 2017 at 155.63 million RUB and declined to 117.24 million RUB in 2019. Revenue values are highly skewed, as seen in the large standard deviations and extreme maximums each year. Average fixed assets remained relatively stable around 110 million RUB until 2018, but dropped significantly to 75.55 million RUB in 2019. The share of firms using the Simplified Accounting Statement (SAS) steadily increased from 48% in 2016 to 56% in 2019.

5 Empirical Strategy

The 2017 reform of the simplified tax system that significantly increased the threshold for company revenues has created a unique environment allowing us to examine the impact of the notch on companies' revenue reporting behavior and as a consequence on their revenue distribution.

Note that the literature exploring individual responses to tax notches observes that a tax notch usually creates local incentives and hence results in local responses just below the threshold. These local responses are typically estimated by the local bunching method (see Kleven (2016) for a survey). However, companies have other margins to respond to a notch and hence their responses may not be purely local. As discussed earlier, companies tend to divide their business to maintain eligibility for the Simplified Tax System (STS). As the model illustrates, such behavior can lead to a redistribution of firm mass not only near the threshold but also at significantly lower revenue levels. Hence, it is necessary to analyze the change

Table 2: Number of Firms by OKVED Sector in 2016-2019

Sector	**2016**	**2017**	**2018**	**2019**
Wholesale and retail trade; repair of vehicles	559 426	545 550	537 527	572 321
Construction	181 172	184 199	190 813	216 477
Manufacturing	153 908	154 729	164 971	175 240
Professional, scientific and technical activities	143 956	144 818	150 050	166 850
Real estate activities	133 875	135 185	141 106	157 459
Transportation and storage	84 739	87 045	90 327	101 281
Administrative and support service activities	71 017	72 051	74 718	82 924
Information and communication	56 873	57 627	59 970	65 720
Accommodation and food service activities	44 359	44 627	46 254	50 977
Agriculture, forestry and fishing	42 092	40 503	40 046	42 665
Other service activities	34 735	33 997	34 418	39 937
Human health and social work activities	31 985	34 058	36 989	41 556
Financial and insurance activities	17 722	15 883	15 296	14 453
Education	15 219	15 217	15 360	19 206
Arts, entertainment and recreation	14 199	14 784	15 969	19 639
Water supply; sewerage, waste management	12 298	12 468	12 743	13 726
Electricity, gas, steam, air conditioning supply	11 896	11 800	11 782	12 222
Mining and quarrying	5 586	5 633	6 630	7 581
Public administration and defence	791	752	1 434	2 030
Activities of households as employers	74	55	69	75
Activities of extraterritorial organizations	4	2	2	3

Table 3: Summary Statistics of Firms for 2016–2019

Year	**2016**	**2017**	**2018**	**2019**
Revenue (mln RUB)				
Mean	140.24	155.63	142.59	117.24
Std. Dev.	19 959.86	31 596.58	8 082.14	5 584.18
Minimum	-199.65	-373.23	-124.53	-42 293.73
Maximum	24 207 453.66	39 149 843.35	6 968 248.04	4 758 711.46
Fixed Assets (mln RUB)				
Mean	104.45	110.48	112.57	75.55
Std. Dev.	12 185.32	12 378.59	12 312.06	11 085.90
Minimum	-38.49	-136.45	-197.06	-106.14
Maximum	7 882 970.56	7 824 129.52	7 864 189.95	7 998 232.55
Share of Firms using Simplified Accounting Statement				
Mean	0.48	0.50	0.52	0.56
Std. Dev.	0.50	0.50	0.50	0.50

in the entire revenue distribution, rather than focusing solely on local bunching near the threshold.

To account for these broader responses, we depart from the standard (local) bunching approach, which estimates excess mass immediately below a notch relative to a locally fitted counterfactual. Instead, we apply the *extended bunching approach* developed by Kosonen and Matikka (2023). Specifically, we estimate how the STS reform affects the broader shape of the revenue distribution, including areas further from the notch. Moreover, since the STS reform shifted the threshold to a substantially higher level, we use the post-reform distribution as the counterfactual for the pre-reform period.

According to the methodology of Kosonen and Matikka (2023), which estimates excess mass across the full distribution below the pre-reform threshold, the bunching coefficient (i.e., excess mass relative to the average density in a bin) is defined as follows:

$$\hat{b}(z) = \frac{\sum_{i=z_L}^{z_N} \left[\frac{c_j^{pre}}{N^{pre}} - \frac{c_j^{post}}{N^{post}} \right]}{\sum_{i=z_L}^{z_N} \left(\frac{c_j^{post}}{N^{post}} \right) / N_B}, \tag{9}$$

where c_j is the number of companies in bin j, and z_i denotes the revenue level in bin j. $\sum_{i=z_L}^{z_N} \frac{c_j}{N}$ reflects the share of companies with the revenue in the range $[z_L; z_N)$, in this case, N is the number of companies in total in the corresponding year. Superscripts "pre" and "post" denotes distributions before and after threshold shifting in 2017. N_B is the number of revenue bins in the revenue range $[z_L; z_N]$. Schematically, the logic of the estimation is presented in Figure 2. In estimation, when we estimate bunching responses to the old threshold, we set the lower limit z_L to 40 mln rubles (the lowest value of revenue in the data) and the higher limit z_N to the old revenue threshold (79,4 mln rubles).

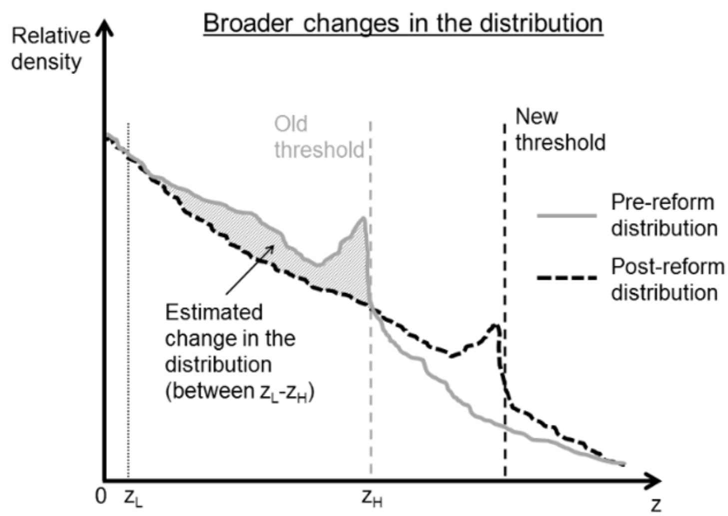


Figure 2: Estimating broader changes in the revenue distribution, source: Kosonen and Matikka (2023)

Next, following Kosonen and Matikka (2023), we incorporate a control group to estimate the causal impact of the STS threshold on the revenue distribution of firms. In general, changes in the revenue distribution may be driven not only by the reform but also by other factors such as macroeconomic fluctuations or broader changes in the business environment. The control group enables us to isolate the effect of the reform by accounting for such confounding factors that might influence revenue patterns independently of the change in the STS threshold.

To implement this, Kosonen and Matikka (2023) propose a modification of Equation 9 that adjusts for concurrent changes in the control group:

$$\hat{b}_d(z) = \left(\frac{\sum_{i=z_L}^{z_N} \left[\frac{c_j^{pre}}{N^{pre}} - \frac{c_j^{post}}{N^{post}} \right]}{\sum_{i=z_L}^{z_N} \left(\frac{c_j^{post}}{N^{post}} \right) / N_B} \right)^T - \left(\frac{\sum_{i=z_L}^{z_N} \left[\frac{c_j^{pre}}{N^{pre}} - \frac{c_j^{post}}{N^{post}} \right]}{\sum_{i=z_L}^{z_N} \left(\frac{c_j^{post}}{N^{post}} \right) / N_B} \right)^C , \tag{10}$$

where superscripts T and C denote treatment and control group. Consequently, this approach allows for a more precise estimation of the changes that are associated with an increase in the revenue threshold, as it additionally considers other potential changes in economic factors and the external environment of companies.

5.1 Splitting into Treatment and Control Groups

The approach outlined in Equation 10 requires a comparison with a control group to isolate the causal impact of the reform. Since our goal is to explore bunching behavior around the revenue threshold for STS eligibility, constructing an appropriate treatment group involves identifying firms that satisfy all STS eligibility criteria except for revenue.

As detailed in Section 2.1, STS eligibility is based on five criteria. For the construction of treatment and control groups, we focus on the four non-revenue criteria: (i) number of employees, (ii) residual value of fixed assets, (iii) ownership structure, and (iv) type of economic activity. Of these, we directly observe only the residual value of fixed assets and industry classification.

However, we also observe whether a firm files a Simplified Accounting Statement (SAS), a method of accounting for financial transactions intended to reduce the accounting burden for small businesses. In Russia, eligibility for SAS is determined by criteria that closely mirror those for the Simplified Tax System (STS), though thresholds may differ. Specifically, to qualify for SAS, firms must have fewer than 100 employees (the same as for STS), annual revenue below 800 million rubles (a higher threshold than for STS), and no more than 25% of equity owned by third-party legal entities (again, the same as for STS).

Since the SAS criteria are either equivalent to or more inclusive than those for STS, all firms eligible for STS are also eligible to file SAS.[9] We exploit this fact to impute missing firm-level characteristics. As our dataset lacks direct measures of employee counts and ownership structure and these two criteria are shared between SAS and STS, we use the SAS indicator as a proxy for fulfilling the STS employee and ownership eligibility conditions. In doing so, we assume that all firms meeting the SAS criteria choose to use it.

Finally, to identify the set of firms eligible for STS, constituting the *treatment group*, we include all firms that filed SAS in a given year and, additionally, control for the type of economic activity (criterion 5) and the residual value of fixed assets (criterion 3).[10] The inclusion of only firms that filed a SAS ensures that eligibility criteria related to the number of employees (criterion 2) and ownership structure (criterion 4) are satisfied.

The *control group* consists of firms that are not eligible for STS based on one or more criteria. As with the treatment group, we use the SAS indicator as a proxy for checking employee and ownership eligibility conditions, and assume that firms file SAS whenever they are eligible to do so. Therefore, firms that do not use SAS are considered not to meet these criteria. However, since SAS eligibility is broader than STS eligibility, we also include in the control group those firms that file SAS

[9]Note that we do not observe the STS status from the data. However, to apply the bunching approach, we only need to know, which firms are eligible for the STS and which are not.

[10]The list of economic activities eligible for STS is provided in Section A of Appendix.

but are not eligible for STS due to their economic activity type or fixed asset value.

Note that the residual-value threshold for fixed assets was also raised in 2017, together with the revenue cutoff. Consequently, the composition of both the treatment group (firms eligible for STS) and the control group (firms ineligible for STS) changes from 2017 onward.

As Kosonen and Matikka (2023) note, the identification assumption is *not* based on random assignment into treatment and control groups. Instead, we assume that the changes in the revenue distribution of the control group reflect the changes in the treatment group in the absence of the reform. In particular, we assume that the relative distributions in the treatment and control groups would have evolved similarly over time if the reform had not occurred.

This identification assumption resembles the parallel trends assumption commonly used in difference-in-differences (DiD) frameworks. To assess its validity, we use additional pre-reform data from earlier years. Figure 3 compares the evolution of the share of firms in the treatment and control groups that have less the certain value of revenue: less than 50, 80, 150, and 180 mln RUB. During the pre-reform period (2012–2015), the trends in these shares are fairly parallel across the two groups, providing support for the parallel trends assumption.

Following the reform, from 2017 onward, we observe that the share of treatment-group firms with revenue below 150 and 180 million RUB increases more rapidly than in the control group. This divergence provides preliminary evidence that firms in the treatment group adjusted their behavior by bunching below the new STS eligibility threshold. In the next section, we formally quantify the magnitude of this bunching response using the extended bunching technique.

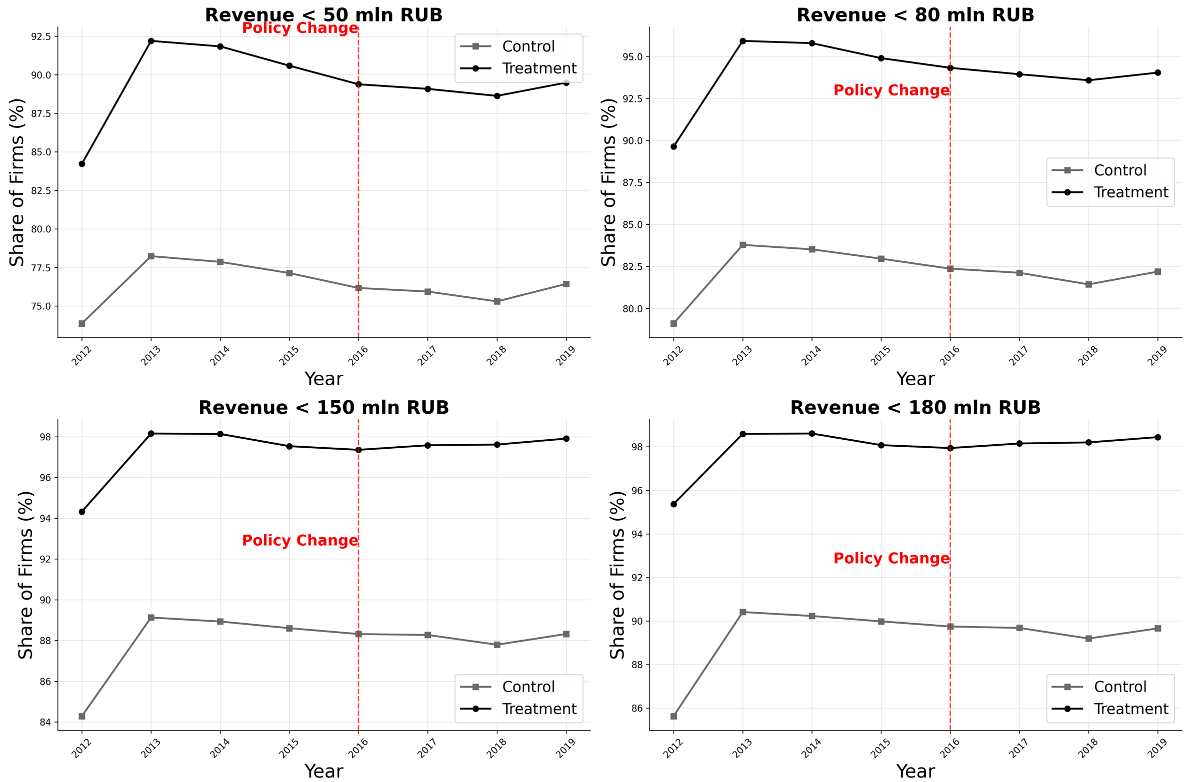


Figure 3: Pre-Trends Comparison for Treatment and Control Groups

Notes: The vertical red dashed line marks the timing of a policy change in 2016. Each panel plots the share of firms for both groups from 2012 to 2019.

6 Results

We begin by examining the distribution of all firms by revenue for the years 2016 to 2019. This provides descriptive evidence of how the revenue distribution evolved around the STS thresholds. We then present the estimation results using the extended bunching method proposed by Kosonen and Matikka (2023), which allows us to quantify the magnitude of firms' behavioral responses to the reform.

6.1 Graphical evidence

Figure 4 shows the distribution of all firms by revenue from 2016 to 2019 and illustrates how this distribution evolved around two key thresholds determining eligibility for the Simplified Tax System (STS): the original threshold at 79.74 million RUB (effective until 2016) and the revised threshold of 150 million RUB (effective from 2017). Complementarily, Figures C.1 in Appendix present enlarged versions of Figure 4, zoomed in around the old and new thresholds respectively.

In 2016, we observe a distinct concentration of firms below the old threshold, which is suggestive of firms' revenue management behavior to remain within the STS. In the years following the reform (2017–2019), this bunching pattern around the old threshold gradually disappears, and the distribution smooths out. By 2019, we observe a mild irregularity just below the new threshold, indicating a delayed response to the policy change. Overall, this figure provides visual evidence of strategic revenue reporting in relation to STS eligibility, particularly prior the reform.

Figure 5 focuses specifically on firms that are eligible to use the Simplified Tax System, forming the treatment group defined earlier. The 2016 distribution exhibits a pronounced drop just above the old threshold, reflecting some bunching as firms sought to stay within the STS limits. After the 2017 reform, which nearly doubled the revenue threshold, bunching near the old threshold largely disappears. By 2018 and 2019, the distribution becomes smoother in this region, although some indication of a new concentration appears just below the updated threshold. These patterns suggest that firms in the treatment group respond to the reform by adjusting their reported revenue, and that behavioral responses shift upward in line with the revised policy. Compared to the full sample in Figure 4, the concentration around the thresholds is more distinct among firms in the treatment group.

Further, the significant shift in the revenue threshold location due to the STS

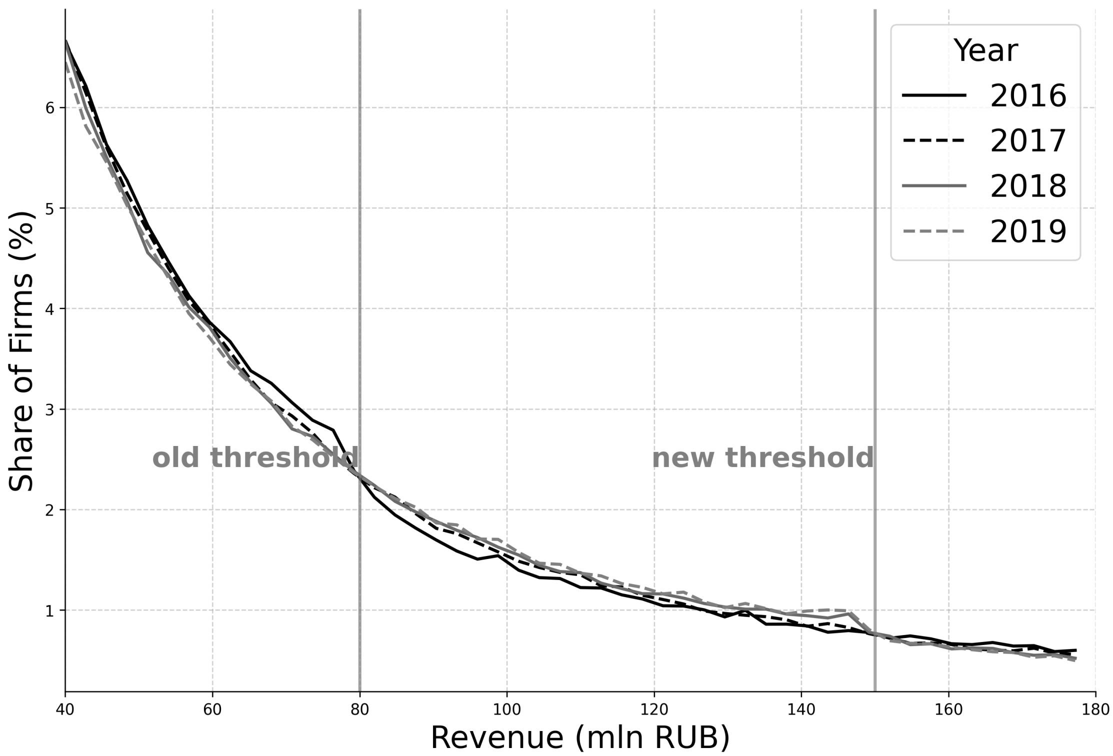


Figure 4: Distribution of all firms by revenue

Note: This figure shows the distribution of firms by revenue for 2016-2019 within a revenue range 40-180 mln RUN in bins of 2.8 mln RUB. The x-axis represents firm revenue in millions of rubles, and the y-axis indicates the share of firms in each revenue bin as a percentage of the total sample. Values are nominal.

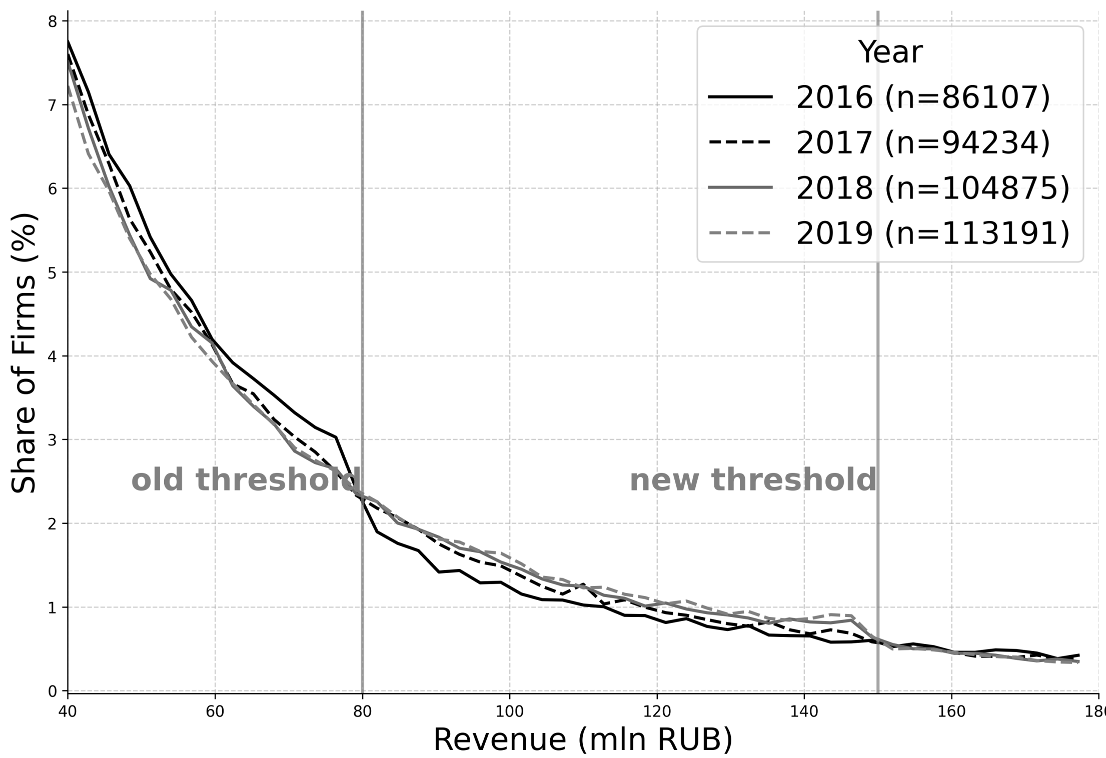


Figure 5: Distribution of firms in the treatment group by revenue

Note: This figure shows the distribution of firms in the treatment group by revenue for 2016-2019 within a revenue range 40-180 mln RUB in bins of 2.8 mln RUB. The x-axis indicates revenue in millions of rubles, and the y-axis shows the share of firms per revenue bin. The legend reports the number of STS firms by year. Values are nominal.

reform allows us to use the after-reform density distribution, i.e., 2017, 2018, or 2019, as a counterfactual distribution for the 2016 pre-reform distribution. Comparing the 2016 distribution to the post-reform years reveals excess mass in the whole distribution below the original threshold. This may suggest that in addition to the visual local bunching responses, there are sizable extensive margin responses of firms through business splitting.

In contrast, Figure 6 presents the revenue distribution for firms not eligible for the STS, our control group. Unlike firms, eligible for the STS, firms in the control group exhibit a smooth distribution across both thresholds throughout the entire period. There is no apparent bunching or discontinuity in the vicinity of either the old or the new thresholds. This stability supports the validity of STS-

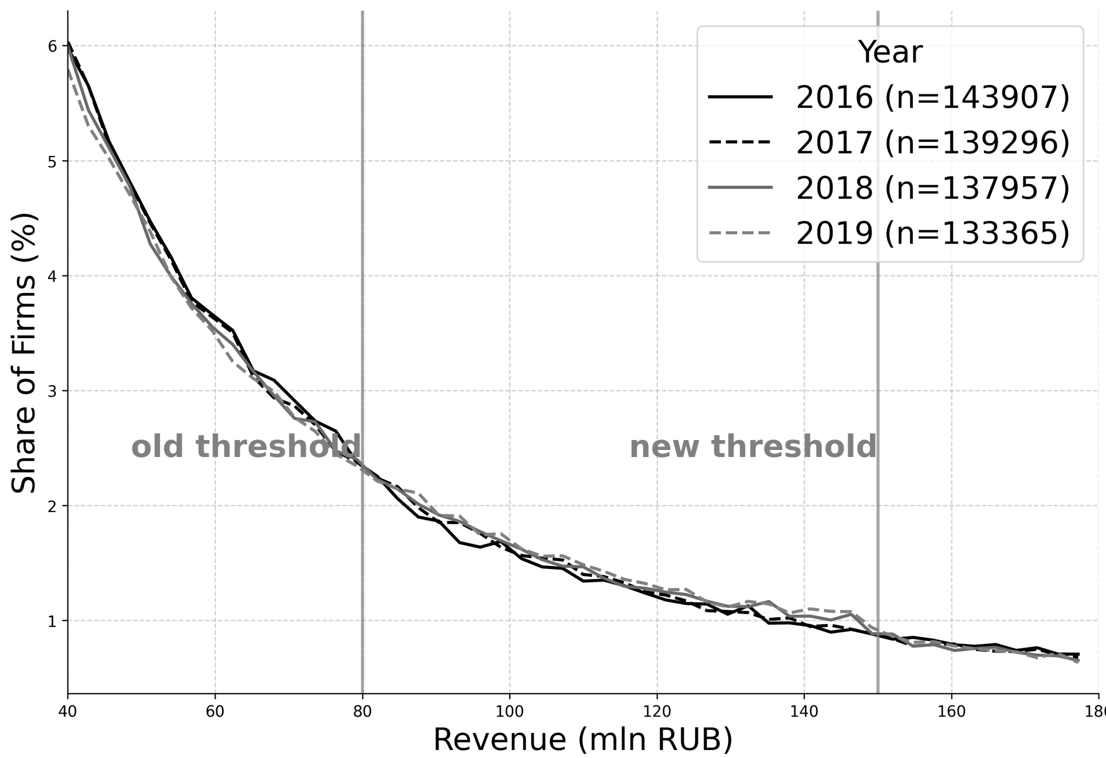


Figure 6: Distribution of firms in the control group by revenue

Note: This figure presents the distribution of firms in the control group for 2016-2019 within a revenue range 40-180 mln RUB in bins of 2.8 mln RUB. The x-axis represents revenue in millions of rubles, and the y-axis denotes the share of firms per bin. The legend indicates the number of firm ineligible for the STS in each year. Values are nominal.

ineligible firms as a control group in analyzing behavioral responses to the revenue thresholds.

To further investigate differential responses to the STS thresholds, Figure 7 presents side-by-side comparisons of revenue distributions for the treatment and control groups in 2016 and 2017.[11] First, note that for the control group, the revenue distributions for 2016 and 2017 mostly coincide and are often indistinguishable. At the same time, for the treatment group, the revenue distributions for 2016 and 2017 are noticeably different. In particular, the treatment group displays a marked concentration of firms (i.e. excess mass) in 2016 compared to 2017 for

[11] Comparisons of revenue distributions for the treatment and control groups in 2018 and 2019 are presented in Figures C.2 and C.3 in Section C of Appendix.

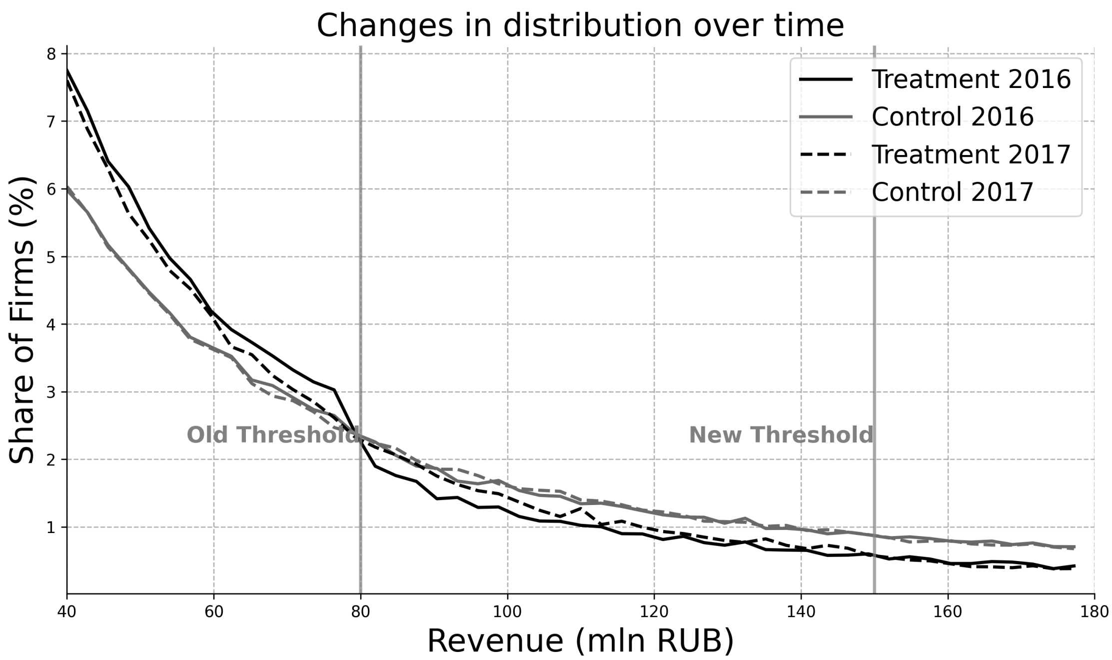


Figure 7: Revenue distribution of treatment and control groups before and after the reform

Note: This figure compares revenue distributions for firms in the treatment and control groups for 2016 (before the reform) and for 2017 (after the reform) within a revenue range 40-180 mln RUB in bins of 2.8 mln RUB. The x-axis represents revenue in millions of rubles, while the y-axis shows the share of firms per bin. Treatment and control groups are plotted in black and gray, respectively. Values are nominal.

revenues below the old threshold, whereas the treatment group distribution in 2016 shows a decrease (i.e. missing mass) compared to 2017 for revenues above the old threshold. Furthermore, a slight uptick in the treatment group's density starts to appear just below the new threshold, indicating an emerging behavioral adjustment to the increased level of the threshold. Note that these behavior responses to the new threshold get larger in 2018 and 2019. Overall, these dynamics highlight how firms eligible for the STS modify their reporting practices in response to policy changes, while firms in the control group remain largely unaffected.

6.2 Estimation Results

6.2.1 Old Threshold

To quantify the strength of bunching responses to the STS notch, we follow the extended bunching approach and estimate Equation 10, which captures the change in the density of treated firms and subtract the change in density observed in the control group. We first evaluate the bunching responses at the old threshold. For this case, the estimation is performed within the revenue range of 40 to 79,7 million RUB. We begin at 40 million RUB to ensure that the range encompasses potential behavioral responses not only near the old threshold (79,7 million RUB) but also further below it.

Table 4 presents estimates of the bunching coefficient at the old threshold in 2016 using the extended bunching method, comparing the revenue distribution in 2016 with that of 2017, 2018 and 2019. The estimated bunching coefficient varies from 1.695 to 2.133 when relying only on the treatment group, as in Equation 9, and from 1.601 to 1.789 when accounting for the change in the control group as in Equation 10. Both estimates are statistically significant. Note that we do not provide a comparison of these estimates with other studies because the bunching coefficient measure we adopted depends on the bin size (i.e., number of bins) , which precludes direct comparisons unless the studies use the same bin size.

Next, we compare these extended bunching estimates to those obtained using the standard local bunching approach estimated over the same revenue interval (Kleven and Waseem, 2013). Methodological details for the local bunching estimation are provided in Section D of the Appendix. Using the local bunching method, we find statistically significant excess mass in all specifications, with estimates ranging from approximately 1.215 to 1.444.

Importantly, the extended bunching point estimates are consistently higher than those from the local approach, and in some specifications, they are statistically

Table 4: Estimates of Excess Mass for Old Threshold

CF Group	Treated		Treated vs. Control	
	Mean	CI	Mean	CI
Extended bunching: comparison with 2017	1.695	[1.472; 1.928]	1.601	[1.333; 1.889]
Extended bunching: comparison with 2018	2.057	[1.840; 2.290]	1.789	[1.492; 2.090]
Extended bunching: comparison with 2019	2.133	[1.921; 2.316]	1.689	[1.409; 1.999]
Local bunching: polynomial of degree 3	1.215	[0.999; 1.455]	-	-
Local bunching: polynomial of degree 4	1.444	[1.186; 1.732]	-	-

Note: The columns labeled *Treated* show the differences for the treatment group. The *Treated vs. Control* columns present the results from a difference-in-differences specification, comparing the change in the treatment group relative to the control group. Confidence intervals are obtained as 2.5% and 97.5% quantiles of the bootstrap distribution with 1,000 replications.

different. This finding aligns with previous evidence in Kosonen and Matikka (2023) and highlights the importance of capturing broader behavioral adjustments beyond the immediate vicinity of the threshold.

6.2.2 New Threshold

Let us now focus more on the reaction of firms to the shift in the threshold. As we have mentioned, while the increase in the threshold occurred from the beginning of 2017, bunching at the new threshold level is just slightly evident in 2017 and becomes sizable only in 2018. This suggests that the adjustment process to such a large shift in the threshold took a year, indicating a delayed reaction of the firms, which might be due to inertia, learning, or lack of knowledge.

Although with a delay, firms eventually began to concentrate at the new threshold. This raises the following question. How large is the new bunching that emerged at the new threshold?

To estimate the excess mass associated with the new threshold, we can now use the 2016 pre-reform distribution as a counterfactual distribution for the post-reform distributions, and hence compare the revenue distributions in 2017, 2018, and 2019 with the distribution in 2016.

However, to do this, we need to check that the point, at which the missing mass due to the old threshold ends, lies reasonably below the new threshold. For this purpose, we need to determine this point (i.e., revenue level) in the 2016 distribution where the missing mass ceases, call it z_M. It corresponds to the revenue level at which the missing mass above the old threshold equals the excess mass below it. That is, this point z_M, can be solved numerically using the following equation:

$$\sum_{i=z_L}^{z_N} \left[\frac{c_j^{pre}}{N^{pre}} - \frac{c_j^{post}}{N^{post}} \right] = - \sum_{i=z_N}^{z_M} \left[\frac{c_j^{pre}}{N^{pre}} - \frac{c_j^{post}}{N^{post}} \right]. \tag{11}$$

Moreover, as Table 4 reports two estimates of the excess mass, one using only the treatment group and another incorporating the control group, we also compute the two corresponding values of z_M. Based on the treatment group alone, z_M is estimated at 132 million RUB. When the control group is accounted for, the estimate increases to 140 million RUB. In doing this, we rely on the comparison of the 2016 distribution with the 2017 distribution.

Table 5: Estimates of Excess Mass for New Threshold

	Treated		Treated vs. Control	
	Mean	CI	Mean	CI
Extended bunching for 2017	1.386	[0.972; 1.816]	0.657	[0.307; 1.043]
Extended bunching for 2018	2.073	[1.672; 2.470]	0.662	[0.315; 1.072]
Extended bunching for 2019	1.862	[1.411; 2.290]	0.799	[0.369; 1.277]

Notes: The columns labeled *Treated* show the differences for the treatment group, where we take 132 mln RUB as the threshold T^*. The *Treated vs. Control* columns present the results from a difference-in-differences specification, comparing the change in the treatment group relative to the control group, where we take 140 mln RUB as the threshold T^*. Confidence intervals are obtained as 2.5% and 97.5% quantiles of the bootstrap distribution with 1,000 replications.

Table 5 reports the estimated excess mass at the new threshold using the extended bunching approach, comparing post-reform years (2017–2019) to the pre-reform distribution in 2016. When analyzing the changes in the treatment group, we observe sizable and statistically significant bunching coefficients, ranging from 1.386 in 2017 to 2.073 in 2018, before slightly declining to 1.862 in 2019. When taking into account the control group as in Equation 10, the estimated bunching coefficient is smaller but remains statistically significant in all years, ranging from 0.657 to 0.799. The lower magnitude reflects the net effect of the reform relative to underlying trends in the control group. Overall, these results suggest a persistent behavioral response to the new STS threshold, with the strongest adjustment occurring three years after the reform.

Note that, given the proximity of the point z_M to the new threshold, we cannot fully estimate the business splitting responses through the extended bunching approach. Therefore, we try to detect some traces of business splitting by looking at the pattern of how firms were born and dissolved over time.[12] Specifically, Figure C.4 in Appendix compares the revenue distributions of three groups of firms in 2018: all existing firms, firms founded in 2017, and firms dissolved in 2019. The y-axis shows the share of firms in each revenue bin relative to the total number of firms in the respective group. The figure shows the following tendencies around the new threshold. Although the firms born in 2017 prevail below the new threshold, the firms dissolved in 2019 prevail above the new threshold. These trends of emergence and disappearance of firms are indicative of business splitting.

7 Conclusion

In this paper, we examine the impact of the Simplified Tax System (STS) in Russia on companies' revenue reporting behavior and, as a consequence, on their revenue distribution. The STS is a size-based regulation and creates a notch in tax respon-

[12] Our dataset lacks information about the ownership structure of the firms, which prevents us from analyzing business splitting using ownership data.

sibilities, because when a company switches from the general tax system to the STS, it pays less taxes and has fewer reporting requirements.

Importantly, the cases of STS tax evasion, uncovered by the Russian tax authority, reveal that in an effort to keep revenues below the STS threshold, firms apply illegal business splitting in addition to other revenue under-reporting practices. The use of business splitting by firms creates a challenge in estimating firms' responses to a size-based regulation. It is because the use of business splitting leads not to a local bunching responses but to reported revenues significantly below the threshold, which makes the existing (local) bunching method not applicable.

To better understand the distortions caused by STS, we develop a model of firms' behavior under a sized-based taxation. In the case where firms evade only by means of under-reporting revenue, which results in intensive margin responses, the size-based taxation leads to (local) bunching at the threshold, occurring through two channels: first, firms reduce their production to an inefficient level, and second, they increase their evasion. In the case where firms have an opportunity to use business splitting, there arises an excess mass of firms with revenue significantly below the threshold, accompanied by a decline in the excess mass of firms that bunch at the threshold. Overall, the mass of firms that illegally qualify for the use of STS increases in this case. This suggests that taking into account business-splitting evasion is important to accurately assess the potential downsides of a size-based regulation, which is crucial to both theoretical and empirical literature.

To estimate the excess mass in the entire distribution below the threshold, our empirical strategy relies on the extended bunching approach recently proposed by Kosonen and Matikka (2022) and utilizes the 2017 reform in the STS rules. This reform significantly increased the revenue threshold, which allows us to use the post-reform revenue distribution as a counterfactual distribution. Additionally, to ensure that we estimate the causal impact of the STS, we incorporate a control group and estimate the bunching coefficient by comparing treated and control firms' distributions before and after the reform.

Using firm-level data for almost all Russian businesses from the recently created Russian Financial Statements Database (RFSD), we construct and analyze the distribution of firms by revenue for 2016 - 2019 years. For the treatment group, the pre-reform distribution, compared to the post-reform distribution, shows excess mass of firm in the whole distribution below the original threshold, indicating the presence of significant business splitting responses in addition to the local bunching responses.

Additionally, we observe that after the threshold was shifted to an increased level, the concentration of firms at the new threshold does not arise immediately but becomes significant only a year after the shift. This indicates a delayed reaction of companies to the large shift in the threshold.

References

ALMUNIA, M. AND D. LOPEZ-RODRIGUEZ (2018): "Under the radar: the effects of monitoring firms on tax compliance," *Am. Econ. J. Econ. Pol.*, 10, 1–38.

BACHAS, P. AND M. SOTO (2021): "Corporate taxation under weak enforcement," *American Economic Journal: Economic Policy*, 13, 36–71.

BEST, M. C., A. BROCKMEYER, H. J. KLEVEN, J. SPINNEWIJN, AND M. WASEEM (2015): "Production versus revenue efficiency with limited tax capacity: theory and evidence from Pakistan," *Journal of Political Economy*, 123, 1311–1355.

BONDARKOV, S., V. LEDENEV, AND D. SKOUGAREVSKIY (2025): "Russian Financial Statements Database: A firm-level collection of the universe of financial statements," *arXiv preprint arXiv:2501.05841*.

CARVALHO, C. C. (2024): "Size-based business taxation in a high-informality context," .

DHARMAPALA, D., J. SLEMROD, AND J. D. WILSON (2011): "Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs," *Journal of Public Economics*, 95, 1036–1047.

DI MARZIO, I., S. MOCETTI, AND E. RUBOLINO (2025): "Market Externalities of Tax Evasion," .

ELSCHNER, C. (2013): "Special tax regimes and the choice of organizational form: Evidence from the European Tonnage Taxes," *Journal of Public Economics*, 97, 206–216.

GALE, W. G., H. GELFOND, A. KRUPKIN, M. J. MAZUR, AND E. J. TODER

(2019): “Effects of the tax cuts and jobs act: A preliminary analysis,” *National Tax Journal*, 71, 589–612.

HARJU, J., T. MATIKKA, AND T. RAUHANEN (2019): “Compliance costs vs. tax incentives: Why do entrepreneurs respond to size-based regulations?” *Journal of Public Economics*, 173, 139–164.

KEEN, M. AND J. MINTZ (2004): “The optimal threshold for a value-added tax,” *Journal of Public Economics*, 88, 559–576.

KLEVEN, H. J. (2016): “Taxes and Bunching,” *Annual Review of Economics*, 8, 1–56.

KLEVEN, H. J. AND M. WASEEM (2013): “Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan,” *The quarterly journal of economics*, 128, 669–723.

KOSONEN, T. AND T. MATIKKA (2023): “Discrete Labor Supply: Quasi-Experimental Evidence and Implications,” .

ONJI, K. (2009): “The response of firms to eligibility thresholds: Evidence from the Japanese value-added tax,” *Journal of Public Economics*, 93, 766–775.

SHARMA, R. R., J. SLEMROD, M. STIMMELMAYR, J. D. WILSON, AND P. CHOI (2025): “Optimal Dual-Regime Business Tax Systems,” .

WEI, F. AND J.-F. WEN (2019): “The optimal turnover threshold and tax rate for SMEs,” .

——— (2023): “Designing a Presumptive Income Tax Based on Turnover in Countries with Large Informal Sectors,” .

ZANONI, W., J. PANTANO, P. CARRILLO-MALDONADO, AND N. CHUQUIMARCA (2025): "Bunching, Tax Regime Notches and Taxpayer Behavior," *Available at SSRN 5068948*.

Appendix

A Types of economic activities for which the use of the simplified tax system is prohibited

According to Article 346.12 of the Tax Code of the Russian Federation, the organizations with the following types of economic activities do not have right to apply for the simplified tax system:

1. Enterprises engaged in banking and insurance activities, including microfinance organizations;
2. Non-state pension funds and investment funds;
3. Professional participants of the stock market;
4. Lombards;
5. Firms that carry out their activities in the production of excisable goods or the extraction and sale of minerals, but common minerals are an exception;
6. Organizations engaged in organizing and conducting gambling activities;
7. Organizations that are parties to production sharing agreements;
8. Organizations and individual entrepreneurs who use the taxation system specifically provided for agricultural producers
9. Treasury and budgetary institutions;
10. Foreign organizations;
11. Private employment agencies engaged in the provision of labor to employees (personnel).

B Proofs

Proof of Lemma 1

a) The FOC $\varphi_y^{'}(y_s^*, a) = 1$, which defines $y_s^*(a)$, corresponds to the solution of the maximization of the before-tax profit, which characterizes the socially efficient level of production. Similar considerations applies to the case of $y_g^*(a)$.

b) Given that $c^{'}(e_g^*) = t_g > t_s$ and $c^{'}(e_s^*) = t_s$, and that $c^{''}(e)$ is positive , it follows that $e_g^* > e_s^*$.

c) Differentiating $\varphi_y^{'}(y_s^*(a), a) = 1$ w.r.t. a, we obtain $\frac{\partial y_s^*(a)}{\partial a} = -\frac{\varphi_{ya}^{''}(y_s^*,a)}{\varphi_{yy}^{''}(y_s^*,a)}$. Given that $\varphi_{yy}^{''}(y,a) > 0$ and $\varphi_{ya}^{''}(y,a) < 0$, we have $\frac{\partial y_s^*(a)}{\partial a} > 0$.

Q.E.D.

Proof of Lemma 2

a) Given that $\varphi_y^{'}(\hat{y}_s, a) = 1 - \frac{\lambda}{1-t_s} < 1$ and $\varphi_y^{'}(y_s^*, a) = 1$, and that $\varphi_{yy}^{''}(y,a)$ is positive, it follows that $\hat{y}_s(a) < y_s^*(a)$.

b) Given that $c^{'}(\hat{e}_s) = t_s + \lambda > t_s$ and $c^{'}(e_s^*) = t_s$, and that $c^{''}(e)$ is positive , it follows that $\hat{e}_s(a) > e_s^*$.

c) Using $\varphi_y^{'}(\hat{y}_s, a) = 1 - \frac{\lambda}{1-t_s}$, $c^{'}(\hat{e}_s) = t_s + \lambda$, $\hat{y}_s - \hat{e}_s = \overline{Y}$, we can obtain $\varphi_y^{'}(\hat{y}_s(a), a) = 1 - \frac{c^{'}(\hat{y}_s(a)-\overline{Y})-t_s}{1-t_s}$. Differentiating this equation w.r.t. a, we obtain
$\frac{\partial \hat{y}_s(a)}{\partial a} = -\frac{\varphi_{ya}^{''}(\hat{y}_s,a)}{\varphi_{yy}^{''}(\hat{y}_s,a)+\frac{1}{1-t_s}c^{''}(\hat{y}_s(a)-\overline{Y})}$.

Given that $\varphi_{ya}^{''}(y,a) < 0$, $\varphi_{yy}^{''}(y,a) > 0$, and $c^{''}(e) > 0$, we have $\frac{\partial \hat{y}_s(a)}{\partial a} > 0$. Additionally, given that $\hat{y}_s(a) - \hat{e}_s(a) = \overline{Y}$ and $\overline{Y}$ is constant, we obtain that $\frac{\partial \hat{e}_s(a)}{\partial a} = \frac{\partial \hat{y}_s(a)}{\partial a} > 0$.

d) First, given that $\hat{e}_s(a)$ increases with a but e_s^* is constant. The difference $\hat{e}_s(a) - e_s^*$ increases with a. Second, given that $\frac{\partial y_s^*(a)}{\partial a} = -\frac{\varphi_{ya}^{''}(y_s^*,a)}{\varphi_{yy}^{''}(y_s^*,a)}$ and $\frac{\partial \hat{y}_s(a)}{\partial a} = -\frac{\varphi_{ya}^{''}(\hat{y}_s,a)}{\varphi_{yy}^{''}(\hat{y}_s,a)+\frac{1}{1-t_s}c^{''}(\hat{y}_s(a)-\overline{Y})}$, we have $\frac{\partial y_s^*(a)-\hat{y}_s(a)}{\partial a} = \frac{-\varphi_{ya}^{''}(y_s^*,a)}{\varphi_{yy}^{''}(y_s^*,a)} - \frac{-\varphi_{ya}^{''}(\hat{y}_s,a)}{\varphi_{yy}^{''}(\hat{y}_s,a)} \frac{1}{(1+\frac{1}{1-t_s}\frac{c^{''}(\hat{y}_s(a)-\overline{Y})}{\varphi_{yy}^{''}(\hat{y}_s,a)})}$.

Because $\frac{-\varphi''_{ya}(y,a)}{\varphi''_{yy}(y,a)}$ is non-decreasing function of y and $y_s^*(a) > \hat{y}_s(a)$, we obtain that $\frac{-\varphi''_{ya}(y_s^*,a)}{\varphi''_{yy}(y_s^*,a)} > \frac{-\varphi''_{ya}(\hat{y}_s,a)}{\varphi''_{yy}(\hat{y}_s,a)}$. Given that additionally $\frac{1}{(1+\frac{1}{1-t_s}\frac{c''(\hat{y}_s(a)-\overline{Y})}{\varphi''_{yy}(\hat{y}_s,a)})} < 1$, we get $\frac{\partial y_s^*(a)-\hat{y}_s(a)}{\partial a} > 0$. Hence, the difference $y_s^*(a)-\hat{y}_s(a)$ increases in a.

Q.E.D.

Proof of Proposition 1

Using the envelope theorem, it is possible to show that $\frac{\partial \hat{\Pi}_s(a)}{\partial a} = (1-t_s)(-\varphi'_a(\hat{y}_s,a))$ and $\frac{\partial \Pi_g^*(a)}{\partial a} = (1-t_g)(-\varphi'_a(y_g^*,a))$. Although $(1-t_g) < (1-t_s)$, we have $y_g^*(a) = y_s^*(a) > \hat{y}_s(a)$ and $-\varphi''_{ay}(y,a) > 0$. Moreover, given that $y_g^*(a)-\hat{y}_s(a)$ increases with a and $-\varphi''_{ay}(y,a) > 0$, there exists $\hat{a}$ such that $\frac{\partial \Pi_g^*(\hat{a})}{\partial a} = (1-t_g)(-\varphi'_a(y_g^*(\hat{a}),\hat{a})) > \frac{\partial \hat{\Pi}_s(a)}{\partial a} = (1-t_s)(-\varphi'_a(\hat{y}_s(\hat{a}),\hat{a}))$.

Q.E.D.

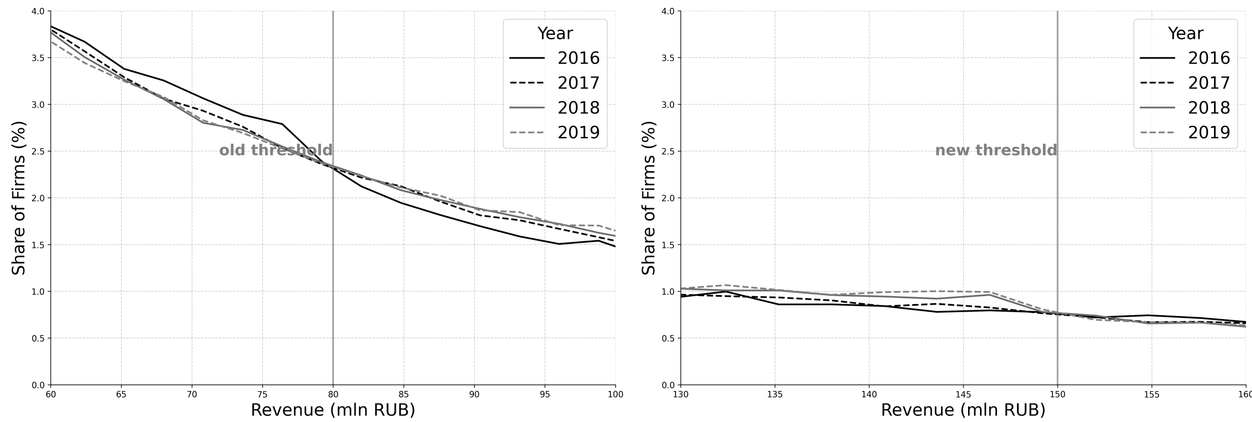
Proof of Proposition 2

At a^*, we have $\Pi_{sb}^{max}(a) = \Pi_{sb}^*(a^*) < \hat{\Pi}_s(a^*) = \Pi_s^*(a^*)$, because $\Pi_{sb}^*(a^*) = \Pi_s^*(a^*) - b$. As productivity a grows, for $a^* \leq a \leq a_b^*$, profit $\Pi_{sb}^*(a)$ grows faster than $\hat{\Pi}_s(a)$ and additionally, for $a > a_b^*$, $\hat{\Pi}_{sb}(a)$ also grows faster than $\hat{\Pi}_s(a)$. Indeed, $\frac{\partial \Pi_{sb}^*(a)}{\partial a} = (1-t_s)(-\varphi'_a(y_s^*,a)) > \frac{\partial \hat{\Pi}_s(a)}{\partial a} = (1-t_s)(-\varphi'_a(\hat{y}_s,a))$, and $\frac{\partial \hat{\Pi}_{sb}(a)}{\partial a} = (1-t_s)(-\varphi'_a(\hat{y}_{sb},a)) > \frac{\partial \hat{\Pi}_s(a)}{\partial a} = (1-t_s)(-\varphi'_a(\hat{y}_s,a))$. Hence, there exists the productivity, called $\underline{a_b}$, when $\Pi_{sb}^{max}(\underline{a_b}) = \hat{\Pi}_s(\underline{a_b})$.

Additionally, because $\underline{a_b} < \hat{a}$, we have $\Pi_{sb}^{max}(a) > \Pi_s^*(a) \geq \Pi_g^*(a)$ for $\underline{a_b} \leq a \leq \hat{a}$. But, as a grows, $\Pi_{sb}^{max}(a)$ becomes equal to $\hat{\Pi}_{sb}(a)$, which grows with $\frac{\partial \hat{\Pi}_{sb}(a)}{\partial a} = (1-t_s)(-\varphi'_a(\hat{y}_{sb}(a),a))$, while $\Pi_g^*(a)$ grows with $\frac{\partial \Pi_g^*(a)}{\partial a} = (1-t_g)(-\varphi'_a(y_g^*(a),a))$. Applying similar arguments as in Proposition 1, we can show that $y_g^*(a)-\hat{y}_{sb}(a)$ increases with a. Using this fact and that $-\varphi''_{ay}(y,a) > 0$, there exists a such that $\frac{\partial \Pi_g^*(a)}{\partial a} > \frac{\partial \hat{\Pi}_{sb}(a)}{\partial a}$. Hence, there exist $\overline{a_b}$ such that $\Pi_{sb}^{max}(\overline{a_b}) = \Pi_g^*(\overline{a_b})$.

Q.E.D.

C Additional Graphical Evidence



(a) Old threshold (b) New threshold

Figure C.1: Distribution of all firms by revenue zoomed for the old and new thresholds

Note: This figure shows a zoomed-in view of the distribution of firms by revenue under the old and the new threshold. The x-axis represents firm revenue in mln RUB, and the y-axis indicates the share of firms in each revenue bin as a percentage of the total sample. Values are nominal.

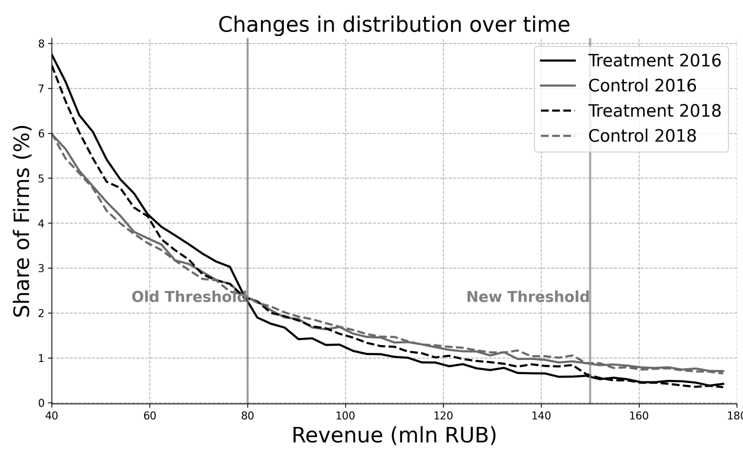


Figure C.2: Distribution of firms by revenue in 2016 and 2018

Note: This figure compares revenue distributions for firms in the treatment and control groups for 2016 (before the reform) and for 2018 (after the reform). The x-axis represents revenue in millions of rubles, while the y-axis shows the share of firms per bin. Treatment and control groups are plotted in black and gray, respectively.

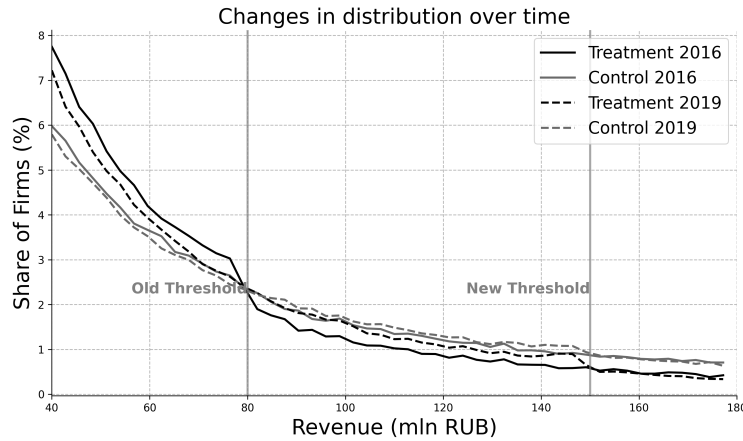


Figure C.3: Distribution of firms by revenue in 2016 and 2019

Note: This figure compares revenue distributions for firms in the treatment and control groups for 2016 (before the reform) and for 2019 (after the reform). The x-axis represents revenue in millions of rubles, while the y-axis shows the share of firms per bin. Treatment and control groups are plotted in black and gray, respectively.

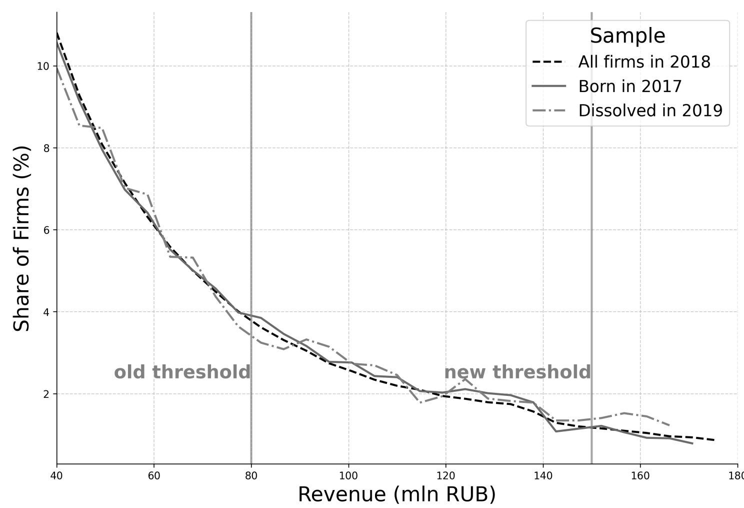


Figure C.4: Comparison of revenue distribution of born, dissolved and all firms in 2018

Note: This figure compares the revenue distributions of three groups of firms in 2018: all existing firms, firms founded in 2017, and firms dissolved in 2019. The x-axis shows revenue (in millions of rubles), and the y-axis shows the share of firms in each revenue bin relative to the total number of firms in the respective group.

D Local Bunching Technique

The standard bunching methodology or *local bunching* is based on fitting a flexible polynomial to the observed distribution (Kleven and Waseem, 2013):

$$c_j = \sum_{i=0}^{p} \beta_i (z_j)^i + \sum_{i=z_{EL}}^{z_{EH}} \eta_i \cdot \mathbf{1}(z_j = i) + \varepsilon_j \tag{12}$$

where c_j represents the count of firms in bin j, and z_j denotes the revenue level associated with bin j. The polynomial order is given by p. We also exclude observations around the threshold from the revenue distribution. Following Kleven and Waseem (2013), the lower boundary z_{EL} of the excluded region is determined visually by identifying the point where bunching behavior becomes evident, i.e., where the density begins to increase. In particular, we exclude observations in the interval from 64 to 96 mln RUB, which corresponds to 20% deviation from the threshold.

The local excess mass near the threshold is estimated by comparing the observed number of firms within the interval (z_L, z^*) to the counterfactual density estimated for the same interval:

$$\hat{b}(z_N) = \frac{\sum_{i=z_{EL}}^{z_N} (c_j - \hat{c}_j)}{\sum_{i=z_{EL}}^{z_N} \hat{c}_j / N_J} \tag{13}$$

where N_J is the number of bins in the interval $[z_{EL}, z_N]$ and z_N is equal to the old revenue threshold (79,4 mln rubles).

To compute standard errors, we follow the bootstrap procedure. Specifically, we generate numerous revenue distributions by randomly resampling firms with replacement. Each resampled distribution yields a new estimate of the counterfactual density. Variation in these bootstrapped estimates is then used to measure uncertainty, with the standard error defined as the standard deviation of the bootstrapped estimates.