FINNISH CENTRE OF EXCELLENCE IN TAX SYSTEMS RESEARCH

FIT Working Paper 50

Sami Jysmä, Youssef Benzarti and Jarkko Harju

Policy Thresholds as Growth Barriers: Theory and Evidence from a Payroll Tax Notch

Tampere University
UNIVERSITY OF HELSINKI
VATT INSTITUTE FOR ECONOMIC RESEARCH
CENTRES OF EXCELLENCE IN RESEARCH

Policy Thresholds as Growth Barriers: Theory and Evidence from a Payroll Tax Notch*

Sami Jysmä† Youssef Benzarti‡ Jarkko Harju§

July 8, 2026

Abstract

Discrete policy thresholds are pervasive in tax and regulatory systems and can substantially distort behavior. We show that notches acting as barriers to mobility within a distribution generate distortions extending far beyond the threshold. The same mechanism biases conventional difference-in-differences estimators, and we propose a new methodology to recover causal effects. Applying the method to the abolition of a size-based payroll tax notch, we find that the notch reduced the number of firms above the threshold by 18 percent and lowered treated firms' employment, capital stock, and value added by 10 percent, whereas conventional difference-in-differences estimates imply negligible effects.

Keywords: Size-based regulation, Notches, Difference-in-differences, Payroll tax
JEL Codes: H30, C01, H25

*We would like to thank David Autor, Raj Chetty, Nathan Hendren, Ohto Kanninen, Antoine Levy, Attila Lindner, Mathilde Muñoz, Emmanuel Saez, and Danny Yagan as well as participants at numerous conferences and seminars for helpful comments. We are grateful to the Research Council of Finland (grant no. 346252 and 364188) for funding for this project. Sami Jysmä is grateful for funding from Palkansaajasäätiö.

†Labore & Tampere University & FIT.

‡University of California at Santa Barbara & NBER.

§University of Glasgow & Finnish Centre of Excellence in Tax Systems Research (FIT) & Tampere University & CESifo & CEPR.

1 Introduction

Discrete policy thresholds are pervasive in tax and regulatory systems. Every OECD country except the United States has a value-added tax, and almost all impose a registration threshold above which firms must charge VAT and bear additional compliance costs (OECD, 2024). A recent OECD review counts more than 2,500 policies targeting small and medium enterprises across member countries, with eligibility typically conditioned on firm size in employment, turnover, or assets (OECD, 2025). How these thresholds shape the firm-size distribution is a central question for both empirical work and policy design. The bunching literature has documented sharp behavioral responses just below such cutoffs and used them to recover behavioral elasticities (Kleven, 2016). We show that these local responses are only part of the story. When a threshold persistently discourages crossing, the effects compound over time and reshape the entire size distribution. The aggregate distortions may substantially exceed those implied by local bunching estimates.

The intuition is simple. A unit deterred from crossing the threshold ends the period smaller than it would otherwise have been. It begins the next period from a lower base, and the shortfall propagates. Across many units and many periods, these missed transitions add up to a reshaped size distribution. Several papers have documented exactly this pattern, in French labor regulations (Garicano et al., 2016; Gourio and Roys, 2014; Aghion et al., 2023) and in French wealth-tax reporting requirements (Garbinti et al., 2023). Studies on the French labor regulations have recovered the cumulative effect by fitting a detailed structural model. This makes it possible to extrapolate to policies beyond what is observed, but the results are hard to transport: each setting requires its own model. The bunching literature (Kleven, 2016) has the opposite problem. It uses behavior in a narrow window below the cutoff to estimate local elasticities and says nothing about firms further out, where most of the cumulative distortion sits. What is missing is a portable framework that quantifies the cumulative effects of policy thresholds without requiring a fully specified model of the underlying process. This paper fills that gap. We show that when a threshold acts as a growth barrier in a stochastic growth process, its distributional effects are summarized by a single sufficient statistic: the probability that a unit refrains from crossing the threshold. This probability can be identified from observed mobility patterns near the cutoff.

We test the framework using the abolition of a firm-level payroll tax notch in Finland. The notch was in place from 1973 to 2009: firms with annual capital depreciation above €50,500 paid an employer payroll tax rate about 2.5 percentage points higher than firms just below it. We use administrative data linking the universe of Finnish firms to their employees and owners from 1996 to 2016. Two findings stand out. First, while the threshold was in force, the mass of firms above the cutoff was approximately 18 percent below its counterfactual level, and the shortfall extended to firms with depreciation more than eight times the threshold. This effect extends far beyond the immediate vicinity of the cutoff and would therefore be largely missed by conventional bunching estimators. The counterfactual distribution implied by our framework, built from pre-reform data alone, closely matches the observed post-reform distribution, providing strong support for the growth-barrier mechanism. Second, applying our switcher difference-in-differences estimator, we find that the notch lowered treated firms' employment, capital stock, and value added by about 10 percent, an effect that conventional difference-in-differences estimates miss almost entirely.

The framework we develop is highly portable. The Finnish payroll tax notch is a fairly generic firm-size threshold: a fiscal cost that turns on at a particular size and has been administratively enforced over decades. Its two key ingredients, a cutoff that persistently discourages crossing and a population whose size evolves stochastically, are also features of VAT registration thresholds, employment cutoffs in labor regulation, or other size-based regulation of firms. With units reinterpreted, the same logic applies to welfare cliffs and the income distribution, or to asset-test thresholds and household saving.

Estimating the causal effect of the threshold on firm-level outcomes turns out to be harder than it looks. Difference-in-differences in its standard form assumes that treatment status is fixed and observed. Policy notches acting as growth barriers violate this assumption. When the threshold is in force, some units stay below it that would otherwise have crossed; when it is removed, those same units begin to cross. These switchers are treated by the policy: they refrained from growing because of it. But a conventional difference-in-differences comparing units below and above the threshold puts them in the wrong group. In the pre-reform period they look like untreated members of the below-threshold control group. Post-reform they appear above the threshold. The control group is contaminated, and the estimator is biased.

Our solution, the switcher difference-in-differences estimator, recovers the average treatment effect on the treated by inferring the share of switchers from the data. When a threshold is removed, the share of units above it rises for two reasons: some switchers cross the threshold, while the rest of the distribution continues to evolve naturally. Parallel trends in counts let us separate the two. With the switcher share identified, the bias in the standard difference-in-differences estimator equals the switcher share multiplied by the pre-reform level gap between the two sides of the threshold. Adding this correction term delivers the average treatment effect on the treated. The procedure needs no data beyond what a standard difference-in-differences analysis uses, and it extends to a multi-period event study framework.

Accounting for switchers is quantitatively important. In our application to the Finnish payroll tax notch, the switcher difference-in-differences estimator shows that the notch reduced treated firms' employment, capital stock, and value added by about 10 percent, while conventional difference-in-differences estimates suggest virtually no response. These differences translate into substantially different implications: we estimate a fiscal externality equal to roughly 2.3 times the mechanical effect of the reform on tax revenues, compared with only 0.1 times under a standard difference-in-differences approach.

The paper contributes to several literatures. The first is the work on behavioral responses to tax notches and regulatory thresholds: turnover and profit-tax cutoffs in Pakistan (Best et al., 2015), tax-enforcement thresholds in Spain (Almunia and Lopez-Rodriguez, 2018), VAT thresholds in Japan and Finland (Onji, 2009; Harju et al., 2019), and size-contingent labor regulations in France (Garicano et al., 2016; Gourio and Roys, 2014; Aghion et al., 2023). Most of this work uses either local responses around the cutoff to identify behavioral elasticities, or a full structural model to estimate the effects of the threshold. We characterize the response further out, where firms operate many times above the threshold, and we estimate effects on firm-level outcomes for the treated firms in a parsimonious difference-in-differences-style framework.

A separate line of work estimates the incidence of payroll taxes on firm scale and worker outcomes (Gruber, 1997; Saez et al., 2019; Cottet, 2023). Our scale and wage results sit comfortably with this evidence. We also extend our earlier paper on the same Finnish reform (Benzarti and Harju, 2021), which used local responses near the threshold to identify firm-level production parameters; the switcher difference-in-differences lets us recover effects for the firms further out that local methods miss. Finally, the theoretical framework builds on the proportional random growth processes long used in the firm-size and wealth-distribution literatures (Gabaix, 2009, 2016; Benhabib and Bisin, 2018), modified to include a policy threshold that acts as a growth barrier.

The switcher difference-in-differences adds to the rapidly growing literature on difference-in-differences (Abadie et al., 2025; Roth et al., 2023; de Chaisemartin and D'Haultfœuille, 2023). Most of this literature addresses challenges arising from staggered treatment adoption or heterogeneity in treatment effects across groups and over time. The switcher problem is different. It arises when a policy reform itself induces some units to cross the boundary that defines the treatment and control groups, contaminating the control group with treated units that cannot be directly identified by the researcher. To address this problem, our bias correction exploits only the observed shares of units on either side of the threshold in each period, allowing us to infer the extent of switching without directly observing the switchers themselves. More broadly, the identification challenge extends well beyond size-based regulation. Similar forms of policy-induced switching arise, for example, when individuals change tax residency across jurisdictions, firms alter their regulatory classification following a reform, or households gain or lose eligibility for welfare programs. In each case, the policy affects group membership itself, creating a source of bias for conventional difference-in-differences estimators.

The rest of the paper proceeds as follows. Section 2 develops the theoretical framework and derives the counterfactual distribution. Section 3 introduces the switcher difference-in-differences. Section 4 covers the Finnish institutional setting, the data, and the empirical results. Section 5 concludes.

2 Theoretical Framework

In this section, we show that size-based policy thresholds, which trigger discrete increases in regulatory burdens, reporting requirements, or fixed costs, can generate substantial distortions in the distributions of affected units, with effects extending well beyond the threshold itself.

For a broad class of growth processes, including the widely used proportional random growth model (also known as geometric Brownian motion; see, e.g., Gabaix 2009), such distortions arise if the threshold reduces the growth rate of units located just below it, for example firms, incomes, or wealth. The key implication is that policy thresholds can act as growth barriers, reshaping the entire distribution instead of merely affecting outcomes near the cutoff. This may have important welfare consequences. For firms, size-based thresholds may reduce aggregate value added and employment; for individuals, they can lower aggregate earnings or wealth.

We first develop intuition using a stylized two-state model. The main insights and results extend to the broader class of growth processes outlined above.

2.1 A stylized two-state model

Consider an economy in which a unit can be either small S^- or large S^+ in a given period. In the absence of policy thresholds, a small unit becomes large with probability q, while a large unit becomes small with probability h. Let F_0^+ denote the steady-state share of large units. In steady state, the flow of units from large to small must equal the flow from small to large: $F_0^+ h = (1 - F_0^+)q$. Solving for F_0^+ and denoting the share of small units as $F_0^- = 1 - F_0^+$, yields the steady state distribution:

$$\begin{cases} F_0^- = \frac{h}{q+h} \\ F_0^+ = \frac{q}{q+h}. \end{cases} \tag{1}$$

Consider next a situation, where a government imposes a fixed cost on large units. We assume that this reduces the probability that a small unit becomes large from q to $(1-p)q$, where p denotes the fraction of units deterred from growing. The new steady-state distribution is then given by:

$$\begin{cases} F_1^- = \frac{h}{(1-p)q+h} \\ F_1^+ = \frac{(1-p)q}{(1-p)q+h}. \end{cases} \tag{2}$$

Suppose that data are available only for periods in which the policy threshold is in place. Hence, we observe the distorted distribution F_1, but not the counterfactual distribution F_0. Our main theoretical result is that identifying the probability p is sufficient to recover F_0. Using Equations (1), (2), and $F_0^- = 1 - F_0^+$ and $F_1^- = 1 - F_1^+$, we obtain the following expression for the counterfactual distribution (in logs):

$$\begin{cases} \log F_0^- = \log(1-p) - \log(1-p(1-F_1^+)) + \log F_1^- \\ \log F_0^+ = -\log(1-p(1-F_1^+)) + \log F_1^+. \end{cases} \tag{3}$$

The intuition behind the result in Equation (3) is that p determines the extent to which the policy threshold reduces the flow of units from the small to the large state. A lower transition rate implies fewer large units and more small units in steady state. Although this result is immediate in the two-state model, the next section demonstrates that the same mechanism operates in considerably more general settings.

2.2 Main model

2.2.1 Setup

We assume that the size S_t of a unit (firm or individual) follows an Itô diffusion in continuous time. Itô diffusion is a continuous stochastic process where the drift and volatility may depend on the current state of the process. In our setting, this means that both growth rates and growth-rate volatility may depend on unit size:

$$dS_t = g(S_t)dt + \sigma(S_t)dW_t, \tag{4}$$

where $g(S_t)$ is the drift, dW_t is a Wiener process,[1] and $\sigma(S_t)$ determines the variance of the process. To ensure integrability, we assume a minimum size at $S_{\min}$, modeled as a reflecting barrier following Gabaix (2009). An alternative approach would be, for example, to introduce a suitable Poisson exit rate for units. The widely used proportional random growth model, or geometric Brownian motion, is a special case of the process in Equation (4), obtained by setting $g(S_t) = gS_t$ and $\sigma(S_t) = \sigma S_t$. Such processes have been extensively used in the literature on firm size distributions because, under mild conditions, they generate power-law distributions that fit observed firm-size distributions well (see e.g. Gabaix 2009 for a review). Equation (4), however, encompasses a broader class of diffusion processes by allowing both drift and volatility to vary with the state of the process. Consequently, it permits relative growth rates and growth-rate volatility to depend on firm size, income, or wealth.

We consider a setting in which the government introduces a policy threshold at size $\bar{S}$, imposing a discrete increase in costs on units above the threshold. We assume that the threshold lowers the growth rate of units located in a small interval below it, $[\bar{S} - \delta, \bar{S}]$, while leaving growth rates elsewhere unchanged. We discuss potential micro-foundations for this mechanism in Section 2.3. This resulting drift function takes the following form:

$$g_1(S_t) = \begin{cases} g(S_t), & \text{if } S_t \notin [\bar{S} - \delta, \bar{S}] \\ g(S_t) - \gamma, & \text{if } S_t \in [\bar{S} - \delta, \bar{S}]. \end{cases} \tag{5}$$

The threshold may induce strategic behavior by creating incentives for units to avoid crossing it. We capture this response by assuming that units located just below the threshold reduce their growth, resulting in a lower drift rate within the interval. Such behavior is plausible in many settings where crossing the threshold triggers substantial fixed costs. In our empirical application, for example, exceeding the payroll tax threshold increases firms' labor costs discretely, providing a strong incentive to remain below the cutoff.

2.2.2 Counterfactual distribution

Next, we use the theoretical framework outlined above to derive an expression for the counterfactual size distribution that would prevail in the absence of the growth barrier. This expression links the counterfactual distribution to two quantities: the observed (distorted) size distribution under the threshold and the probability p that a unit refrains from crossing the threshold due to the policy notch.

We show, with a detailed derivation provided in Appendix B, that for an Itô diffusion of

[1] A Wiener process is a continuous-time stochastic process with zero drift and Gaussian increments. It can be seen as the continuous-time limit of the discrete-time random walk process.

the form given in Equation (4), subject to a growth barrier as specified in Equation (5), the counterfactual log distribution can be expressed as follows:

$$\log f_0(S) = \begin{cases} \log(1-p) - \log(1-p(1-F_1^+)) + \log f_1(S), & \text{if } S \in [S_{\min}, \bar{S}-\delta[\\ \log(1-p) - \log(1-p(1-F_1^+)) + \frac{2\gamma}{\sigma^2}[S-(\bar{S}-\delta)] + \log f_1(S), & \text{if } S \in [\bar{S}-\delta, \bar{S}] \\ -\log(1-p(1-F_1^+)) + \log f_1(S), & \text{if } S \in]\bar{S}, \infty], \end{cases} \tag{6}$$

where p is the probability that a unit refrains from crossing the threshold with the policy in place, $f_0(s)$ is the counterfactual pdf of the distribution, $f_1(s)$ is the observed pdf with the threshold in place, and $F_1^+ = \int_{\bar{s}}^{\infty} f_1(s)ds$ denotes the observed share of units above the threshold.

When δ is small, an estimate of $\log f_0(S)$ can be obtained using estimates of p and the moments of the observed distribution $f_1(S)$:[2]

$$\widehat{\log f_0}(S) = \begin{cases} \log(1-\hat{p}) - \log\left(1-\hat{p}(1-F_1^+)\right) + \log f_1(S), & \text{if } S < \bar{S} \\ -\log\left(1-\hat{p}(1-F_1^+)\right) + \log f_1(S), & \text{if } S > \bar{S}. \end{cases} \tag{7}$$

The intuition for this result is straightforward. By discouraging units from crossing the threshold, the growth barrier reduces the mass of units above $\bar{S}$ and increases the mass below it.[3] Consequently, relative to the counterfactual distribution, the observed distribution is shifted upward below the threshold and downward above it. Importantly, the theory predicts that these shifts are level shifts for log distributions. Equation (7) implies that the difference between these shifts in log densities is $\log(1-p)$.

Figure 1a illustrates these shifts in the context of a firm distribution. The solid black line shows the observed log density $\log f_1(S)$, while the dashed black line shows the counterfactual density $\log f_0(S)$. The shaded red area between $\bar{S}-\delta$ and $\bar{S}$ represents the low-growth region induced by the policy threshold. The blue shaded area denotes the observed mass of firms above the threshold, F_1^+

2.3 Microfoundations

The mechanical framework above does not explicitly model the behavior of the units. Instead, it assumes that the drift of the size process is reduced in a neighborhood below the threshold. Next, we sketch a continuous-time model in which this drift reduction arises endogenously

[2] We provide a method of estimating p in Section 4.3.

[3] Our analysis focuses exclusively on intensive margin responses and growth effects, omitting extensive margin responses, such as firm entry and exit.

from optimal firm behavior. A full treatment is provided in Appendix C. Note that we first abstract from instantaneous bunching, which generates only a local distortion in the model. However, further below we extend the framework to incorporate static bunching alongside the dynamic growth responses.

The model provides one example of microfoundations for the mechanical framework. Appropriate microfoundations may differ across applications; what matters for the applicability of the framework is that the policy threshold acts as a growth barrier.

Intuitively, the option to remain below the threshold and avoid its fixed cost makes additional growth less valuable for firms that are already close to the cutoff, so these firms optimally slow their growth. This concentrates a reduction in drift in a band just below the threshold, reproducing the growth barrier assumed in the mechanical framework.

We assume that firm size S_t follows a controlled Itô diffusion:

$$dS_t = \big(g(S_t) + a_t\big)\,dt + \sigma(S_t)\,dW_t, \qquad a_t \geq 0,$$

where a_t denotes costly growth effort that increases expected growth at a convex flow cost $c(a)$. For firms, this effort may take the form of searching for new business opportunities, pursuing investments, or engaging in innovation activity.

Firms earn flow profits $\pi(S)$ and face a regulatory notch at $\bar{S}$ that imposes a fixed compliance cost $F > 0$ whenever the firm operates above the threshold:

$$\Pi(S) = \pi(S) - F\mathbf{1}\{S \geq \bar{S}\}.$$

Firms choose growth effort a_t to maximize discounted firm profits:

$$V(S) = \max_{\{a_t \geq 0\}} \mathbb{E}_S\left[\int_0^\infty e^{-\rho t}\Big(\Pi(S_t) - c(a_t)\Big)\,dt\right],$$

where $\rho > 0$ denotes the discount rate.

The notch affects growth incentives through the marginal value of firm size. When S is far below $\bar{S}$, an incremental increase in size has little effect on the likelihood of paying the fixed cost F, so optimal effort is close to the no-notch benchmark. As S approaches $\bar{S}$ from below, however, additional growth increases the probability of entering the costly regime and therefore raises the expected discounted cost of the notch. This reduces the marginal value of firm size, $V'(S)$, and consequently lowers optimal growth effort.

To connect this mechanism to the reduced-form framework, we define the endogenous drift as

$$g_j(S) \equiv g(S) + a_j^*(S), \qquad j \in \{0, 1\},$$

where $a_0^*(S)$ denotes the optimal effort in the absence of the notch and $a_1^*(S)$ denotes optimal effort when the notch is in place. The resulting reduction in growth rates is

$$\gamma(S) \equiv g_0(S) - g_1(S) = a_0^*(S) - a_1^*(S) \geq 0.$$

Under mild regularity conditions[4], $\gamma(S)$ is concentrated near $\bar{S}$ and decays quickly away from the threshold. Therefore, the endogenous response can be approximated by a constant reduction in drift over a narrow interval below the threshold:

$$g_1(S) \approx \begin{cases} g_0(S), & S < \bar{S} - \delta, \\ g_0(S) - \gamma, & S \in [\bar{S} - \delta, \bar{S}], \\ g_0(S), & S > \bar{S}, \end{cases} \tag{8}$$

which corresponds directly to the growth barrier in our mechanical framework.

Note however, that our mechanical model assumes that the drift distortion is exactly zero above $\bar{S}$. Instead, the modeled microfoundations generate a local deviation from this just above the threshold, because a firm just above $\bar{S}$ may find it optimal to shrink back below to avoid paying F. However, this reinforces the growth barrier mechanism. Consequently, abstracting from this above-threshold distortion is conservative and incorporating it would imply even larger effects of the threshold.

Incorporating Static Bunching into the Dynamic Framework The microfoundations above treat the size measure S itself as the diffusing state variable. As a result, the notch slows growth near the threshold but does not generate the discrete bunching at $\bar{S}$ typically associated with notched policies. Here we show that standard static bunching can be layered on top of the dynamic mechanism without altering the main logic of the model. In particular, bunching at $\bar{S}$ and the growth barrier summarized by p can coexist within a single framework, arising through distinct channels, and thus they are complementary rather than competing mechanisms. This extension is important as many applications – including our own – exhibit bunching in the context of notches.

Suppose that the Itô diffusion governs an underlying state (e.g. total factor productivity)

[4] See details from Appendix C.

A_t instead of firm size S_t itself, so that

$$dA_t = \big(g(A_t) + a_t\big)\,dt + \sigma(A_t)\,dW_t, \qquad a_t \geq 0,$$

and firms choose S_t (e.g. the capital stock or labor force) statically at each instant:

$$S_t^*(A_t, F) = \arg\max_S \big\{\pi(A_t, S_t) - F \cdot \mathbf{1}\{S_t \geq \bar{S}\}\big\}.$$

Here $S_t^*(A_t, F)$ denotes flow profits, which are assumed to be strictly concave in S_t. with $\partial^2\pi/\partial A_t \partial S_t \geq 0$. Then the frictionless optimal size $S^*(A_t, 0)$ is strictly increasing in productivity A_t.

Let A^L denote the productivity at which the frictionless optimal size reached the threshold, so that $S^*(A^L, 0) = \bar{S}$. Similarly, let A^U denote the productivity level at which a firm is indifferent between locating at the threshold and paying the compliance cost:

$$\pi\big(A^U, S^*(A^U, F)\big) - F = \pi\big(A^U, \bar{S}\big).$$

The resulting static choice of size S_t given A_t is similar to standard bunching literature on notches (e.g. Kleven and Waseem (2013)), firms with $A_t \in [A^L, A^U]$ optimally bunch at the threshold $\bar{S}$. Firms with $A_t > A^U$ choose their frictionless size $S^*(A_t, 0)$, as the fixed cost no longer distorts the intensive margin. Consequently, no firms locate in the interval $(\bar{S}, S_0^*(A^U, F))$, generating the familiar excess mass at the threshold and missing mass immediately above it.

The dynamic problem is unchanged except that the flow payoff becomes

$$\tilde{\pi}(A_t) \equiv \pi(A_t, S^*(A_t, F)) - F \cdot \mathbf{1}\{S^*(A_t, F) \geq \bar{S}\}.$$

Relative to the frictionless benchmark, the notch has no effect for firms below A^L. For firms in the bunching region $[A^L, A^U]$, the cost of the notch increases continuously from 0 to F. For firms above A^U, the firm pays the full compliance cost F, but the choice of S_t is again undistorted. Thus, the discrete notch in observed size S_t becomes a smooth reduction in payoffs over the interval $[A^L, A^U]$.

The analysis of the previous subsections then applies directly. The reduction in payoffs lowers the marginal value of productivity in the vicinity of the bunching interval, reducing optimal growth effort that materializes to lower growth rates in that region. Far from the interval, incentives are approximately unaffected. The sufficient statistic p retains its interpretation as the probability that a firm refrains from crossing the size threshold $\bar{S}$, which occurs when paying the compliance cost becomes optimal (i.e., at productivity level A^U).

The observed size distribution therefore combines two distinct distortions. First, the static choice of S_t generates the familiar bunching pattern: excess mass at $\bar{S}$ and missing mass on $(\bar{S}, S^*(A^U, 0))$. Second, the dynamic growth barrier generates broader shifts in the distribution by reducing drift of firms near the threshold. Bunching therefore creates a local distortion at the cutoff, whereas the growth barrier generates global shifts in the distribution.

3 Switcher Difference-in-Differences

The theoretical framework in Section 2 demonstrates that policy thresholds can act as growth barriers, reshaping entire distributions instead of merely affecting behavior near the cutoff. This insight motivates our empirical approach but also creates a methodological challenge. When a threshold is removed, some units that were previously deterred from crossing it may relocate across the threshold, making it difficult to identify which units were "treated" by the policy. This section develops the switcher difference-in-differences estimator to address this identification challenge.

3.1 An identification issue

When a policy notch is introduced or removed, which units are actually treated? Standard empirical strategies such as difference-in-differences assume that the treatment status is observed for each unit. If this assumption held in our application, observing whether a firm is above or below the threshold would be sufficient to determine its treatment status. However, notched policies can alter the distribution of units by inducing some of them to cross the threshold. These switchers are affected by the policy, yet they appear below the threshold prior to the reform and are therefore classified as untreated by conventional methods. As a result, the control group is contaminated, leading to biased estimates.[5]

To address this, we develop a method that recovers treatment effects using only observed changes in outcomes and unit counts on either side of the threshold. The key insight is that, under the assumptions outlined below, the number of switchers and the resulting bias can be inferred from the observed data.

[5] A similar issue arises for Gruber-Saez-style instruments (Gruber and Saez, 2002). Conventional IV approaches also require observing the actual treatment status of each unit to construct a valid instrument.

3.2 Setup and assumptions

We consider a setting with two periods: a pre-reform period $t = 0$ in which the policy notch is in place and a post-reform period $t = 1$ in which the notch has been repealed. The key challenge is that treatment status is not directly observed. When the notch is removed, some units that were previously deterred from crossing the threshold relocate above it. These units are affected by the policy, yet they are classified as untreated by conventional difference-in-differences methods because they are observed below the threshold before the reform.

Let $y_{f,t}$ denote the realized outcome of unit f in period t. In each period, units belong to one of three latent groups: *below-stayers* (bs), which remain below the threshold regardless of the policy; *above-stayers* (as), which remain above the threshold regardless of the policy; and *switchers* (sw), which locate below the threshold when the notch is in place but above it when the notch is removed.[6] These groups are illustrated in Figure 1b.

Switchers and above-stayers are treated by the reform, as they would locate above the threshold in the absence of the notch. Below-stayers are untreated. When the notch is repealed, switchers move from below the threshold to above it, reducing the number of units below the threshold and increasing the number above. This movement is the source of the bias in conventional difference-in-differences estimators.

The number of units in group $g \in \{bs, as, sw\}$ is denoted by n_t^g, and the average outcome of the group is $y_t^g \equiv E(y_{f,t}|f \in g \text{ during } t)$. Using potential outcome notation, let $y_t^g(R) = E(y_f(R)|f \in g \text{ during } t)$ denote the potential expected outcome of group g in period t when the notch is either still in place ($R = 0$) or repealed ($R = 1$). Finally, let $n_{a,t}$ and $y_{a,t}$ denote the observed number and average outcome of units above ($a = 1$) or below ($a = 0$) the threshold in period t.

We make the following identifying assumptions:

Assumption 1 (Parallel trends in log numbers of units)**.**

$$\log n_1^{bs} - \log n_0^{bs} = \log n_1^{as} - \log n_0^{as} = \log n_1^{sw} - \log n_0^{sw}.$$

This assumption requires that the number of below-stayers, above-stayers, and switchers grow at the same rate. Equivalently, the population shares of the three groups remain constant over time. It implies that, in the absence of the reform, the log number of firms above versus below the threshold would have evolved in parallel. The plausibility of the assumption can be assessed by examining whether the numbers of units above and below the threshold exhibit parallel trends in the pre-reform period. As with standard difference-in-differences designs, however, the assumption is ultimately untestable.

[6] A given unit may belong to a different latent group in different periods.

Assumption 2 (Parallel trends in mean outcomes)**.**

$$y_1^{bs}(0) - y_0^{bs}(0) = y_1^{as}(0) - y_0^{as}(0) = y_1^{sw}(0) - y_0^{sw}(0).$$

This is a standard additive parallel trends assumption applied to the latent groups. It requires that, in the absence of the reform, the mean outcomes of below-stayers, above-stayers, and switchers would have developed similarly over time. Since switchers remain below the threshold when the reform does not occur, the validity of this assumption can be assessed using pre-reform outcome trends.

Assumption 3 (Potential outcomes of switchers without the reform)**.**

$$y_t^{sw}(0) = y_t^{bs}.$$

This assumption states that switchers have the same mean outcomes as below-stayers while they remain below the notch. It allows the observed below-stayer mean to proxy for the unobserved counterfactual outcomes of switchers. The assumption is most plausible when switchers are drawn broadly from the below-threshold population rather than from a narrow bunching region near the cutoff. In our empirical application, we observe some bunching below the threshold, but its magnitude is modest relative to the substantial distributional changes induced by the reform. Moreover, if the distribution of outcomes below the threshold stays similar pre- versus post-reform (up to a constant shift for time trends), this assumption is arguably more plausible. This is the case as large deviations from this assumption are likely to change the outcome distribution below the threshold, given that the outcome distribution of switchers would have been different from the outcome distribution of the below-stayers. However, the assumption is ultimately untestable.

Assumption 4 (Below-stayers not affected by the reform)**.**

$$y_t^{bs}(0) = y_t^{bs}(1).$$

This assumption is a variant of the stable unit treatment value assumption (SUTVA). It requires that the reform does not affect the outcomes of below-stayers. Unlike the standard SUTVA, it does not require the absence of spillovers among treated units. Rather, it only requires that any spillovers do not affect the untreated group. In our application, this assumption is supported by the absence of meaningful changes in outcomes below the threshold following the reform. It could be violated if the reform generated substantial general-equilibrium effects, for example by inducing expansion among above-threshold firms that altered labor market conditions for firms below the threshold.

3.3 Expression for ATET

Our causal parameter of interest is the average treatment effect on the treated (ATET), defined as

$$\text{ATET} = E\big(y_{f,1}(1) \mid f \text{ is treated}\big) - E\big(y_{f,1}(0) \mid f \text{ is treated}\big), \tag{9}$$

where $y_{f,t}(R)$ denotes the potential outcome of unit f in period t under reform status $R \in \{0,1\}$, with $R = 1$ corresponding to the post-reform regime.

We show in Appendix D that, under Assumptions 1–4, the ATET is identified by

$$\text{ATET} = \underbrace{\tau^{DID}}_{\text{Diff-in-diff estimate}} + \underbrace{w(y_{1,0} - y_{0,0})}_{\text{Switcher term}}, \tag{10}$$

where $\tau^{DID} = (y_{1,1} - y_{0,1}) - (y_{1,0} - y_{0,0})$ is the standard repeated cross-sections difference-in-differences estimate comparing means post- and pre-reform below and above the notch point, and $w = 1 - \frac{F_1^+}{F_0^+}$ is the share of switchers among treated units in the post-reform period, where F_{1-R}^+ denotes the share of units above the threshold either when the notch is in place ($R = 0$) or has been repealed ($R = 1$). F_{1-R}^+ is the same object as in Section 2.

Equation (10) provides a simple decomposition of the treatment effect. The first term, τ^{DID}, is the conventional repeated cross-sectional difference-in-differences estimate. The second term corrects for the contamination of the control group switchers among treated units, w, and the pre-reform outcome gap between units above and below the threshold $(y_{1,0} - y_{0,0})$.

If there are no switchers, then $w = 0$ and the difference-in-differences estimate is unbiased. When switchers are present, the correction term is generally non-zero. In particular, if outcomes are on average higher above the threshold than below ($y_{1,0} > y_{0,0}$), the standard difference-in-differences estimator understates the true ATET.

The intuition is straightforward. Conventional difference-in-differences treats all units observed below the threshold before the reform as untreated. However, some of these units are switchers who would have located above the threshold in the absence of the notch and are therefore treated by the policy. Their counterfactual outcome equals the below-stayer mean, which is lower than the counterfactual outcome of above-stayers. By failing to account for these lower counterfactual outcomes, the standard estimator overstates the counterfactual mean of the treated group and consequently understates the treatment effect. The magnitude of this bias increases with both the share of switchers and the outcome gap across the threshold.

3.4 Regression implementation

Equation (10) admits a transparent regression interpretation. Consider the standard repeated cross-sectional difference-in-differences specification:

$$y_{f,t} = \alpha + \beta A_f + \gamma P_t + \delta(A_f \times P_t) + \varepsilon_{f,t}, \tag{11}$$

where $A_f = \mathbf{1}(f \text{ is above the threshold in period } t)$ and $P_t = \mathbf{1}(t \text{ is post-reform})$. In the 2×2 cell means, $\hat{\delta} = \tau^{DID}$ and $\hat{\beta} = y_{1,0} - y_{0,0}$. Hence, the ATET can be expressed in terms of the regression coefficients:

$$\text{ATET} = \hat{\delta} + \hat{w}\hat{\beta},$$

where $\hat{w}$ is computed from observed counts via $\hat{w} = w = 1 - \frac{F_1^+}{F_0^+}$. The bias of the standard DiD regression estimator $\hat{\delta}$ equals $-\hat{w}\hat{\beta}$: the product of the switcher share and the pre-reform level gap between the above- and below-threshold groups.

Since $\hat{w}$ is estimated from changes in the distribution of observed units rather than from outcome variation, the ATET cannot be recovered as a single coefficient from a standard OLS regression. We therefore propose the following two-step procedure: (i) estimate equation (11) by OLS and obtain $\hat{\delta}$ and $\hat{\beta}$, (ii) estimate the switcher share $\hat{w}$ from the observed counts above and below the threshold. Then the treatment effect is $\widehat{\text{ATET}} = \hat{\delta} + \hat{w}\hat{\beta}$. Standard errors can be obtained by bootstrapping the entire procedure and resampling units with replacement.

3.5 Event study

The switcher difference-in-differences estimator extends naturally to a multi-period event-study design. Suppose we observe T periods $t \in \{1, \dots, T\}$, with the reform taking effect after period t^*. Consider the event-study regression:

$$y_{f,t} = \alpha A_f + \sum_t \gamma_t D_t + \sum_{t \neq \bar{t}} \delta_t (A_f \times D_t) + \varepsilon_{f,t}, \tag{12}$$

where $A_f = \mathbf{1}(f \text{ is above the threshold in period } t)$, D_t are time dummies, and the interaction coefficient for a reference pre-reform period $\bar{t} \leq t^*$ is normalized to zero. The OLS coefficient $\hat{\alpha} = y_{1,\bar{t}} - y_{0,\bar{t}}$ captures the outcome gap between units above and below the threshold in the reference period, while $\hat{\delta}_t$ is the conventional event-study estimate for period t relative to $\bar{t}$.

The corrected treatment effect for each period $t \neq \bar{t}$ is then

$$\widetilde{\text{ATET}}_t = \hat{\delta}_t + \hat{w}_t \hat{\alpha}, \tag{13}$$

where $\hat{w}_t$ denotes the share of switchers among treated units in period t.

Plotting $\widetilde{\text{ATET}}_t$ together with the uncorrected estimates $\hat{\delta}_t$ provides a transparent visualization of switcher bias over time. The difference between the two series equals the correction term $\hat{w}_t \hat{\alpha}$. The correction may vary across periods if adjustment to the reform is gradual, causing the share of switchers among treated units to evolve over time.

3.6 Estimating the switcher term from pre-period data only

In many applications, the researcher only observes the data from periods when the policy threshold is in place. In these settings, the researcher can still estimate the switcher term from Equation (10) relying on the diffusion theory of Section 2. Hence, the tradeoff is between data requirements and additional structure.

To see this, note that the switcher term is

$$w(y_{1,0} - y_{0,0}),$$

where $y_{1,0}$ and $y_{0,0}$ are the pre-period expected values of the outcome above and below the threshold.

Hence, the question is how to identify w, which is the share of switchers among treated units after the threshold is repealed. As noted above, this share can be expressed as $w = 1 - \frac{F_1^+}{F_0^+}$, where F_{1-R}^+ is the share of units above the threshold either when the threshold is in place ($R = 0$) or not ($R = 1$). Hence, if we have data only from periods with the threshold ($R = 0$), the only unidentified moment is the counterfactual share of units above the threshold F_0^+. However, using Equation (7) from our diffusion theory, we can express $F_0^+ = \frac{1}{1-p(1-F_1^+)} F_1^+$, where p is the probability that firms refrain from growing past the threshold because of the notch. Applying this expression, we then have that

$$w = p(1 - F_1^+), . \tag{14}$$

Hence, we can obtain a theory-based estimate of the switcher term as

$$\widehat{w(y_{1,0} - y_{0,0})} = \hat{p}(1 - \hat{F}_1^+)(\hat{y}_{1,0} - \hat{y}_{0,0}), \tag{15}$$

with an estimate of p.[7]

4 Empirical Application: A Finnish Payroll Tax Notch

This section tests the theoretical predictions from Section 2 and applies the estimation methodology outlined in Section 3 to an empirical application. We begin by briefly describing the Finnish payroll tax system, which featured a size-based threshold that created a tax notch for firms above a certain cutoff. Then we describe the data used to implement the empirical framework developed above. Finally, we present and discuss the results, focusing on their implications for a range of firm-level outcomes.

4.1 Institutions

Finland operates a comprehensive public social insurance system, which covers pensions, unemployment insurance, accident insurance, health insurance, and life insurance. These programs are financed through payroll taxes collected from both employers and employees. Employers contribute a larger share, accounting for roughly two thirds of total payroll tax payments. In 2023, the average employer payroll tax rate was approximately 17% of monthly gross wages, while the employee rate was about 8%.

Both employer and employee characteristics determine payroll tax contribution rates. For most workers, the employee payroll tax rate is 7.15%, but for employees aged 53–62 the rate is approximately 1.5 percentage points higher than for younger or older workers. This paper focuses on a reform that abolished an employer-side size-based threshold, above which the employer payroll tax rate increased by about 2.5 percentage points, accounting for roughly 15% of total firm-level payroll taxes. The threshold was defined in terms of annual firm-level capital depreciation and was set at €50,500, above which the higher rate applied. The threshold was calculated based on the previous year's capital depreciation.

Moreover, the tax incentive at the threshold varied by firms' capital intensity. The relevant measure was the ratio of capital depreciation to labor costs. Firms with capital depreciation below 10% of labor costs faced no discontinuity at the threshold and thus had no incentive to adjust. Firms with capital depreciation between 10% and 30% of labor costs experienced an increase of roughly 2 percentage points in their payroll tax rate at the threshold. Finally, firms with capital depreciation exceeding 30% of labor costs faced, on average, a 2.5 percentage point increase in their payroll tax rate. Appendix A, Table A1

[7] We provide a way of estimating p in Section 4.3.

summarizes these categories, while Appendix A, Table A2 reports all variation in payroll tax rates across years and firm categories during our study period.

This stepwise increase in payroll taxes by firm-level capital depreciation was introduced in Finland in 1973 (HE, 1972). It formed part of a broader social security reform related to pensions and housing benefits, which were projected to raise government expenditures. To ensure budget neutrality, the government increased payroll taxes for a targeted group of firms, focusing on those that were relatively more capital intensive (in relation to labor costs), while exempting many self-employed and small businesses whose payroll tax rates remained unchanged.[8]

Following an agreement by the employer and employee unions in January 2009, the Finnish Government made a proposal to repeal the threshold in September 2009 (HE, 2009). The Parliament decided to enact the repeal in November 2009, and the repeal took effect in January 2010. As the threshold was calculated based on the previous year's capital depreciation, we consider 2009 to have been the year of the reform. After the reform, all firms, regardless of whether their capital depreciation was below or above €50,500, faced the same payroll tax rate. At the same time, a small reduction in the overall employer payroll tax rate was implemented, applying uniformly to firms on both sides of the former threshold.

Suitability of the Notch for Testing Theory. This notch provides a suitable setting for testing the predictions of the theory and for applying the switcher DiD design developed above. The policy created a discontinuous increase in labor costs that depended on the firm-level capital through annual depreciation. As a result, labor costs were not directly affected through marginal adjustments in wages or employment, but rather through the amount of capital held by the firm.

That said, the institutional setting is not perfectly aligned with the theoretical framework. The notch may also influence marginal incentives through the firm-level production function. For instance, if firms were considering expanding their scale by increasing both capital and labor prior to the reform, crossing the threshold would imply a slightly higher marginal cost of labor as well. However, this change in marginal incentives is small, approximately a 2.5 percent increase, and is therefore unlikely to be the primary driver of our results.

[8]From 1973 to 1981 the threshold was set at 20,000 Finnish Marks. Between 1982 and 1987 it was raised to 50,000 Finnish Marks, and from 1988 onward it remained nominally fixed at 300,000 Finnish Marks until 2001. In 2002, with Finland's entry into the euro area, the threshold was converted to €50,500 at the standard exchange rate. The only exception was the period between January 1992 and June 1993, when payroll taxes were not linked to capital depreciation.

4.2 Data

Firm-level data. We use comprehensive data on all firms operating in Finland between 1996 and 2016. The dataset combines tax records with firms' financial accounts. The tax data include all items used to compute corporate taxes, such as net-of-VAT revenue, variable costs, capital depreciation, number of employees, labor costs, taxable profits, and dividends. The accounting data contain the financial variables required by international accounting standards, including revenue, value added, and accounting-based measures of capital depreciation. In addition, we observe firms' industry code, organizational form, location, and age. Our analysis primarily relies on firm-level information, although we also use employee- and owner-level outcomes, which we discuss in more detail below.

Job-spell data. Our job-spell data contain information on all employees, which we can merge with the employer-level data using firm-level identifiers. This employee-level data are reported at the end of each year but they include all firm-level job contracts at the daily level with the starting and ending dates for each employee-firm job contract pair. This dataset also includes unique employee-level identifiers. We use this data from 1996 to 2016.

Individual-level data. We observe all individual-level annual income information and a large set of background characteristics. This dataset also contains individual identifiers which we can use to merge into the data sets we have described above. The income variables include earnings, capital income, disposable income and government subsidies. The background characteristics include the standard set of demographic variables such as age, gender, education, and place of residence. These data are available from 1987 onward until 2016 for all individuals living in Finland on the last calendar date of each year.

Owner-level data. We can link all the owners of corporations, partnerships and sole proprietors to the firm-level dataset where we have both firm- and owner-level identifiers. These data include the ownership share, type of owner (firm, individual, foreign owners) and the number of owners. These owner-level data also have a unique identifier which we can use to merge with the background characteristics dataset. This dataset is only available from 2006 onward, limiting our analysis to those years when we examine owner-level outcomes.

Sample restrictions and modifications. We winsorize observations with extreme outcome values, above the 95th or below the 5th percentile of each examined continuous outcome variable at a yearly level. All of our continuous monetary variables are shown in 2010 prices, normalized using the consumer price index.

Table 1 presents the descriptive statistics of all firms in our sample in the first column, for our baseline estimation sample between 100 log points below and above the threshold (€18,578 and €137,273 of annual depreciation) and for the last 20 log points of firms just below the threshold (between €41,348 and €50,499 of annual depreciation) in 2005–2007.

Finally, Figure 2a plots the variation we exploit: firm-level payroll tax rates, in percentage points, across the firm-level capital depreciation distribution, as described in Section 4.1. The bin size is defined such that each bin corresponds to 20 log points of capital depreciation, and this same bin definition is used consistently throughout the paper for firm distributions over capital depreciation. The drop in the overall payroll tax burden above the threshold is approximately 2.5 percentage points, which corresponds to a 12 log points decrease in the average firm-level payroll tax rate.

This institutional setting provides an ideal context for applying our framework: the sharp threshold created an incentive for firms not to grow past the threshold, the reform removed it completely, and the comprehensive administrative data allow us to observe firm-level responses.

4.3 Estimating p and the Counterfactual Distribution

According to the theoretical framework developed in Section 2, the threshold should first constrain firm growth by reducing the probability p that a firm crosses it. Second, if the growth barrier mechanism is indeed relevant, the theoretical counterfactual firm distribution derived using the estimated value of p should closely approximate the observed firm distribution after the threshold was repealed. To test this and provide evidence for the validity of our approach, we construct the theoretical counterfactual distribution and compare it to the post-reform distribution observed after the abolition of the tax notch.

To estimate p, we define the treatment group as the 60 log points of depreciation closest to, but below, the threshold. This group is selected because it exhibits significant bunching in growth rates near the threshold[9]. The control group consists of the 60 log points further below the threshold. The exact depreciation values in the treatment group range from €27,715 to €50,499, while those in the control group range from €15,210 to €27,714.[10]

Following Garbinti et al. (2023), we define the normalized growth rate for the treatment group as

[9] We show this bunching later in Figure 2c

[10] Note that the treatment group here is different from the baseline treatment group in our difference-in-differences setup in Section 4.4. We do this as p is a local parameter that applies only to firms that would have grown past the threshold (and thus are fairly close to the threshold to begin with), while the ATET estimates in the DiD setup aim to capture the full effect of the reform.

$$\bar{g}^1_{i,t} = \frac{d_{i,t+1} - d_{i,t}}{d_{i,t}} - \frac{50500 - d_{i,t}}{d_{i,t}} = \frac{d_{i,t+1} - 50500}{d_{i,t}}, \tag{16}$$

which represents the growth rate of firm i's capital depreciation $d_{i,t}$, net of the growth that would have brought the firm to the threshold at €50,500.

We adopt a data-driven approach to scale the growth-rate distribution of the control group to construct a counterfactual distribution for the treatment group. More specifically, we define the counterfactual normalized growth rates of the control group as

$$\bar{g}^0_{i,t}(\hat{\alpha}) = \frac{d_{i,t+1} - \hat{\alpha} \times 50500}{d_{i,t}}, \tag{17}$$

where $\hat{\alpha}$ is chosen to minimize the distance between selected moments of the distributions of $\bar{g}^1_{i,t}$ $(h_1(g))$ and $\bar{g}^0_{i,t}(\hat{\alpha})$ $(h^0_1(g|\hat{\alpha}))$. The objective function for this minimization is the mean squared error between the cumulative distribution functions (CDFs) of these two distributions, computed up to a given level of normalized growth g^*.[11] Ideally, g^* should be selected such that there is no evident bunching below that point. In our baseline specification, we set $g^* = -0.04$; however, in our setting, the choice of this parameter has little impact[12]. In other words, $\hat{\alpha}$ is chosen by

$$\hat{\alpha} \equiv \arg\min_{\alpha} \frac{1}{N_b} \sum_{b \in [b_{\min}, g^*]} [H_1(b) - H^0_1(b|\alpha)]^2, \tag{18}$$

where N_b is the number of normalized growth rate bins below g^*, $b_{\min}$ is the lowest normalized growth rate bin, and $H_1(b)$ and $H^0_1(b|\alpha)$ are the empirical CDFs of normalized growth rates $\bar{g}^1_{i,t}$ and $\bar{g}^0_{i,t}(\alpha)$.

Figure 2 shows the distributions of $\bar{g}^1_{i,t}$ for the treatment group and $\bar{g}^0_{i,t}(\hat{\alpha} = 0.535)$ for the control group. Panel 2c shows the PDFs of these distributions and 2d shows the corresponding CDFs. In both panels, the growth-rate bunching of the treatment group is clearly visible just below the threshold. This pattern indicates that a substantial share of firms that would have grown beyond the threshold refrain from doing so when the threshold is in place. In other words, the threshold acts as a growth barrier. To quantify this share, we estimate the CDF of the treatment and control groups just below the threshold and compute the corresponding probability as

$$\hat{p} = \frac{\text{Bunching mass}}{\text{Counterfactual mass above threshold}} = \frac{H_1(0) - H^0_1(0|\hat{\alpha} = 0.535)}{1 - H^0_1(0|\hat{\alpha} = 0.535)}. \tag{19}$$

[11] CDF is a natural target here as our estimate of p is a function of the CDFs of the treatment and control group.

[12] Choosing g^* between 0 and -0.2 did not change the choice of $\hat{\alpha}$ at the level of three decimal places.

Here, the numerator $H_1(0) - H_1^0(0|\hat{\alpha} = 0.535)$ represents the bunching mass below the threshold. This extra mass corresponds to firms that would have crossed the threshold in the absence of the barrier. The numerator is divided by the total mass of firms that would have crossed the threshold had it not existed, which is given by the mass above the threshold in the counterfactual distribution $(1 - H_1^0(0|\hat{\alpha} = 0.535))$. Using this framework, we obtain an estimate of $\hat{p} = 0.182$ with a bootstrapped standard error of 0.001.

Counterfactual and Before-After Firm Distributions. Using the estimate of $\hat{p}$ and the empirical share of firms above the threshold in the absence of the policy, F_1^+, we construct the counterfactual firm distribution. Figure 2b plots this counterfactual distribution (dashed black line, log PDF) for 2005–2007, together with the observed distributions with the threshold in place (years 2005–2007, blue solid line) and without the threshold (years 2011–2013, red solid line).[13]

The figure shows that, during 2005–2007, the firm-size distribution exhibits bunching just below the threshold, and notably, there are substantially fewer firms far to the right of it. This pattern already suggests sizable firm-level responses to the payroll tax rate increase. Another key finding is that the theoretical counterfactual distribution aligns closely with the observed distribution in the post-reform period 2011–2013. The counterfactual distribution was estimated using only data from 2005–2007. Moreover, we do not impose any functional form on the firm distribution, for example, to capture a level shift at the threshold. Instead, we rely solely on the estimated growth-barrier probability p and the share of firms above the threshold F_1^+ from the 2005–2007 data. We then adjust the observed firm distribution from that period according to shifts suggested by our theoretical results in Section 2. The close fit between the counterfactual and post-reform distributions therefore provides strong support that the threshold acted as a growth barrier, with potentially large implications for the entire firm distribution.

Figure 2b also illustrates that firm distributions by depreciation follow a power law[14], appearing approximately as a straight line in log-log space (see, e.g., Gabaix 2016) after the repeal of the threshold in 2011–2013. In contrast, when the threshold was in place in

[13]In this analysis, we exclude the three-year period from 2008 to 2010, as the reform was first publicly discussed in the media on 10 January, 2009. However, because representatives of both employee and employer organizations were involved in drafting the legislation, some firms may have anticipated the reform as early as 2008.

[14]A power law is defined as a relationship between two variables, Y and X, such that $Y = aX^{\beta}$, where β is the so-called power law exponent and a is a constant. When X, for example, increases by 10, Y increases by 10^{β}, implying that Y changes with X according to the power of β. Without the notch, our estimate of β is -0.710 (0.017), which is highly statistically significant and relatively close to the value reported by Gabaix 2016, who finds $\beta = -1.069$, when using the number of employees as a proxy for firm size for large firms in the United States.

2005–2007, the firm distribution can be characterized as following a split power law: while slope of the distribution above the threshold remains similar to that below it, the level is noticeably lower.[15] [16]

Robustness of the estimated p. As the choice of the control group in the dynamic bunching estimation is somewhat arbitrary, we assess the robustness of the estimated p to alternative control group definitions. In our baseline counterfactual definition, the control group consists of firms located 60–120 log points below the threshold, corresponding to annual depreciation between €15,210 and €27,714. We then consider two alternative control groups, comprising firms located 120–180 and 180–240 log points below the threshold, corresponding to depreciation levels between €8,347 and €15,210, and between €4,581 and €8,347, respectively. These two alternative definitions yield estimates $\hat{p} = 0.200$ and $\hat{p} = 0.257$, respectively, both with bootstrapped standard errors of 0.001. Appendix A, Figure A4, plots the corresponding counterfactual distributions for the three control group definitions while holding the remaining aspects of the estimation unchanged. The estimated counterfactual distributions are highly similar across specifications, indicating that the results are not particularly sensitive to the choice of control group.

Evolution of the Firm Distribution. Appendix A Figure A5 plots the log probability density function (PDF) of firms by annual depreciation for six selected years spanning the period before and after the abolition of the payroll tax notch. The vertical line marks the notch threshold. Prior to the reform (years 2005 and 2007), the firm distribution is more concentrated below the threshold. Following the repeal, the distribution gradually shifts to the right and the mass of firms above the threshold increases steadily, and by 2015 the

[15] Appendix A Figure A2 further examines the changes in firm distributions using alternative definitions. Panel (a) of Figure A2 shows the log of firm frequency by bins on the vertical axis against the log of firm-level capital depreciation in the same bins on the horizontal axis, comparing periods with the notch (2005–2007) and without the notch (2011–2013). Panel (b) of Figure A2 plots the corresponding firm density distributions before and after the repeal of the threshold, illustrating that the impact extends beyond the local area around the threshold – even when the number of firms is held constant across periods. Panels (c) and (d) of Figure A2 plot the log frequency and density of firms, respectively, using a balanced panel of firms in both periods (2005–2007 and 2011–2013). These panels reveal a substantial shift in the firm distribution, implying significant movement of firms within the distribution.

[16] Appendix Figure A3 further investigates firm movement across the capital depreciation distribution before and after the payroll tax discontinuity. Panel (a) of Figure A3 shows that a large share of firms remained in the same depreciation bin just below the threshold prior to 2010, indicating clear bunching behavior when the payroll tax rate increased at the threshold. This local behavioral response disappears after the tax notch is abolished. Panels (b) and (c) demonstrate that the threshold constrained regular firm mobility within the distribution, as reflected by marked decreases in both measures. Panel (d) of Figure A3 shows that average exit rates decline above the threshold in the post-2010 period (without the notch) compared with the pre-2010 period (with the notch). We define a firm as exiting if it ceases to file tax reports in the subsequent year and never reappears in our sample thereafter.

distribution has shifted upward over a broad range of firm sizes. This pattern suggests that the notch acted as a persistent barrier to firm growth, generating a shortage of larger firms while it was in place. Consistent with our theoretical framework, these changes extend well beyond the immediate vicinity of the threshold.

Real Responses versus Avoidance. A potential explanation for the observed firm distribution is that firms engage in tax avoidance or evasion strategies to remain below the threshold, rather than genuinely reducing their scale or growth. Such responses could include manipulating accounting measures, splitting firms into smaller entities, or substituting owned capital with rented capital or outsourced activities. However, the evidence presented in Appendix E suggests that these mechanisms are unlikely to account for our findings. Specifically, we show that accounting-based depreciation, which is subject to third-party auditing for almost all firms in our sample, exhibits distributional responses similar to those based on tax depreciation. This indicates that our results are not driven by tax accounting practices. We further show that firm splits cannot explain the observed patterns, and that neither rental capital nor outsourcing appears to be an important margin of adjustment. In sum, these findings suggest that the estimated distributional responses primarily reflect real changes in firm growth rather than avoidance or evasion behavior.

4.4 Difference-in-Differences Results

Next, we present results on firm performance and scale using the switcher difference-in-differences (DiD) framework introduced in Section 3, together with the institutional setting and data described above. In our baseline setup, the treatment group consists of above-stayers and switchers with capital depreciation within 100 log points above the threshold (capital depreciation between €50,500 and €137,273) in the absence of the notch. The baseline control group includes below-stayers with capital depreciation within 100 log points below the threshold (capital depreciation between €18,578 and €50,499).[17]

We begin by examining a set of standard firm-level outcomes that are directly affected by the payroll tax increase at the threshold, using equations (12) and (13). Panel (a) of Figure 3 presents two dynamic event-study estimates over time for labor costs: our newly developed *Switcher DiD*, and a standard repeated cross-section DiD (*RCS DiD*)[18] Both estimators

[17]Note that the treatment group here is different from the baseline treatment group in our dynamic bunching setup in Section 4.3. We do this as p is a local parameter that applies only to firms that would have grown past the threshold (and thus are fairly close to the threshold to begin with), while the ATET estimates in the DiD setup aim to capture the full effect of the reform.

[18]Event-studies using a firm fixed effect and treatment assignment based on whether a firm was above (treated) or below (control) the threshold in 2007 are shown in Appendix Figures A6 and A7. However,

exhibit stable pre-trends prior to the 2009 reform, supporting the validity of the parallel trends for mean outcomes assumption.

At the time of the reform, however, the *Switcher DiD* estimates start to increase markedly, while the *RCS DiD* estimates continue to follow patterns similar to their pre-reform trends. When pooling the estimates over the study window to before (2005–2007) and after (2011–2013) periods, the pooled *Switcher DiD* implies an economically and statistically significant increase of 14 log points in labor costs relative to the control group (Table 2 summarizes the pooled point estimates for all baseline outcome variables). In contrast, the pooled *RCS DiD* estimates indicate virtually no response, with a coefficient of 0.004 (SE 0.021). This stark difference highlights the importance of accounting for the distributional shift through the correction term in our empirical setting.

Panel (b) of Figure 3 reports estimates for average firm-level log wages (net of payroll taxes), a commonly used measure of tax incidence in the payroll tax literature (see, e.g., Gruber 1997). The figure shows stable pre-trend estimates tightly centered around zero across both estimators. While the *RCS DiD* estimate indicates no response, our newly developed *Switcher DiD* estimator implies an increase in average wages of 0.034 (SE 0.008) following the reform (see Table 2). This suggests that the reduction in payroll taxes translated into higher wages within firms, with an effect roughly comparable in magnitude to the tax cut. However, in a setting characterized by substantial movement in the firm distribution, these estimates are difficult to interpret strictly as measures of tax pass-through.

One caveat is that this wage measure does not account for compositional changes in the workforce. Potential changes in the characteristics of newly hired or separated workers before and after the reform could influence the observed wage patterns. However, Appendix Figure A8 presents estimates of the shares of non-tertiary-educated and routine workers among firms in the treatment and control groups using the same empirical framework, and shows no immediate or medium-term changes in these measures. Overall, the results indicate no substantial compositional shifts in the workforce. That said, the share of routine workers declines slightly following the reform, which may contribute to the observed increase in wages.

Consistent with the evidence above, Panel (c) of Figure 3 shows that the increase in labor costs documented in panel (a) is mainly driven by employment responses. While pre-trends remain stable across estimators, the *Switcher DiD* estimates reveal a sharp increase in average employment at the time of the 2009 reform. The pooled estimate in Table 2 indicates an increase of approximately 12 log points in employment among treated firms relative to the control group. In contrast, the *RCS DiD* estimator shows little to no response.

---

these estimates suffer from severe mean reversion as can be seen from the event-studies.

Panel (d) of Figure 3 provides evidence that the increase in firm scale through employment is accompanied by a sizable and comparable rise in capital, as measured by fixed assets. The bias in the repeated cross-section difference-in-differences estimator is equally pronounced for capital stock as it is for labor costs and employment in panels (a) and (c).

Panels (a)–(c) of Figure 4 show that the scale effects are also consistently evident in firms' revenue, value added and profits after the reform. The *Switcher DiD* estimates in Table 2 suggest systematic size effects on all of these three outcomes with pooled point estimates ranging from 10 to 13 log points. Also, similarly as above, these outcomes show a similar systematic bias in the repeated cross-section difference-in-differences estimates.

Finally, Panel (d) of Figure 4 reports labor share estimates over time (the wage bill divided by value added). The pooled estimate suggests a small positive effect of 0.008 (SE 0.004), which is equivalent to a 2.1 percent increase from the average labor share in our treatment group of 0.377 in 2007.

Taken together, overall labor costs increase significantly among firms above the threshold after the abolition of the tax notch, which is mainly driven by an increase in number of employees, although we also find an increase in average wages among firms affected by the reform. This sizable increase in labor costs and employment is accompanied by comparable growth in other firm scale measures. We observe similar effects for capital stock, revenue, value added, and profits, all suggesting that firm size significantly increased in response to the payroll tax cut.

Assessing the Switcher Difference-in-Differences Assumptions. We next assess the plausibility of the identifying assumptions underlying the switcher difference-in-differences estimator. These assumptions are introduced in Section 3.

We begin with the parallel trends in log counts assumption (Assumption 1), which requires that, absent the reform, the (log) number of firms above and below the notch threshold would have evolved in parallel. As in a standard difference-in-differences setting, the plausibility of this assumption can be evaluated by examining the pre-reform trends. Appendix Figure A9 shows the evolution of the difference in log counts above and below the threshold[19]. This difference remains stable throughout the pre-treatment years but increases sharply after the repeal of the notch in 2009, supporting the parallel trends in log counts assumption.

Next, we consider the parallel trends in mean outcomes (Assumption 2), which requires that, in the absence of the reform, average outcomes above and below the notch would have evolved similarly over time. We can assess the validity of this assumption using a standard

[19] Appendix Figure A9 is for firms with observed log labor costs. The corresponding figures for other outcomes are similar, as the only difference is whether a given outcome is observed for a particular firm.

difference-in-differences approach. The repeated cross-section analysis presented in Figures 3 and 4 allows us to do this. The pre-trends for our main outcomes are relatively stable, supporting the validity of this assumption.

Finally, we assess the assumption that switchers and below-stayers have the same average outcome prior to treatment (Assumption 3). We study the evolution of the outcome distribution below the threshold before and after the reform. In the pre-reform years, both below-stayers and switchers are located below the threshold, whereas after the reform only below-stayers remain. Thus, systematic differences between the two groups would imply changes in the below-threshold outcome distribution over time. To abstract from aggregate trends, we de-mean the pre- and post-distributions. This exercise does not constitute a formal test if switchers and below-stayers differ only by an additive constant, but this scenario appears unlikely in our setting, where the main source of heterogeneity is expected to arise from bunching. Appendix Figures A10 and A11 present kernel density estimates of the demeaned outcome distributions below the threshold before and after the reform. The distributions are generally similar, supporting the plausibility of the assumption, although moderate shifts are visible for log revenue, log employment, and the labor share.

Theory Based Estimates of the Switcher Term. As discussed in Section 3.6, the theoretical framework developed in Section 2 can be combined with pre-reform data to construct ex ante estimates of the switcher term. These theory-based estimates for our main outcomes are reported in the bottom panel of Table 2. Across outcomes, the theory-based estimates are slightly larger than the switcher term estimates obtained using both pre- and post-reform data. However, the two approaches yield estimates of a similar order of magnitude. For labor costs, number of employees, capital stock, revenue, value added, and profits, the estimated switcher terms exceed 10 log points, whereas the corresponding estimates for log average wages are between 3 and 4 log points, and those for the labor share are slightly below 1 percentage point. The theory-based estimates of the switcher share more clearly exceed those obtained from the empirical data.

Robustness of the Difference-in-Differences Results. As the definition of the treatment and control group is somewhat arbitrary, we assess the robustness of the switcher difference-in-differences results to these choices. We first estimate the ATET for alternative choices of the donut hole around the threshold and then for alternative bandwidths, which determine how far from the threshold firms are included in the estimation sample.

Appendix Figures A12 and A13 show estimates for varying sizes of donut holes for the baseline outcomes, while keeping all other aspects of the specification fixed. Our baseline estimates without a donut hole are indicated by red triangles. Overall, the baseline estimates are representative of those obtained across most donut-hole definitions, although the estimates for log labor costs, log average wages, and log number of employees are slightly larger in the baseline specification than in specifications excluding firms close to the threshold.

Appendix Figures A14 and A15 present estimates for alternative bandwidth choices. Our baseline specification, which uses a bandwidth of 200 log points, is again indicated by red triangles. The estimated effects are broadly stable across bandwidth choices, although the estimates for log capital stock, log revenue, log value added, and log profits tend to increase with the bandwidth. A likely explanation for this pattern is that the pre-reform outcome gap between firms above and below the threshold becomes larger as the estimation window expands, increasing the magnitude of the switcher correction for these outcomes.

4.5 Aggregate and Fiscal Implications

Given the sizable firm-level estimates above, it is natural to ask what the reform implied for the economy as a whole. We consider two questions: how much the repeal changed aggregate quantities, and how much it altered government revenues once behavioral responses are accounted for. Appendix F describes the methodology used to construct these estimates in detail.

The estimated firm-level effects are substantial, amounting to roughly 10 log points for employment, labor costs, and value added. However, these effects apply only to firms affected by the notch, which account for a relatively small share of the aggregate economic activity. Aggregating the estimated responses across the full firm distribution, we find that the repeal of the notch raised aggregate outcomes ranging from approximately 1.5 percent for capital stock to 5 percent for labor costs. Thus, while the aggregate effects are smaller than the corresponding firm-level estimates, they remain economically meaningful given the relatively narrow reach of the reform.

The reform also had fiscal consequences that extend well beyond its direct mechanical effect. Holding firm behavior fixed, repealing the notch reduced payroll tax revenues collected from firms previously subject to the higher tax rate. However, the induced growth in firm scale broadened several tax bases simultaneously. Higher employment and wages increased payroll and earned income tax revenues, higher value added increased value-added tax revenues, and higher profits increased corporate income tax revenues. Summing across these behavioral margins, we estimate a fiscal externality equal to approximately 2.3 times the mechanical revenue loss, implying that the additional revenues generated by firm growth more than offset the direct fiscal cost of the reform.

By contrast, a conventional analysis based on standard repeated cross-section difference-in-differences estimates implies a fiscal externality of only about 0.1 times the mechanical revenue effect, substantially understating the revenue feedback. As with the firm-scale estimates above, this discrepancy reflects switcher bias: conventional estimators attribute little or no growth response to the reform and therefore fail to capture most of the additional revenues generated by switcher firms.

5 Conclusion

This paper develops a theoretical framework, a new empirical methodology, and provides evidence on the effects of discrete policy thresholds using the abolition of a firm-level payroll tax notch in Finland. Our analysis makes three main contributions.

First, we show theoretically that when policy notches act as growth barriers, they generate distortions that extend far beyond the threshold itself. For a broad class of stochastic processes governing firm or individual dynamics, such barriers systematically reduce mass above the threshold while increasing mass below it. Crucially, these distributional effects can be characterized using a simple sufficient statistic: the probability that a unit refrains from crossing the threshold when the notch is in place. This probability can be quasi-experimentally identified using the dynamic bunching approach of Garbinti et al. (2023), and it allows us to construct a counterfactual distribution that would prevail in the absence of the threshold. While our application focuses on firms, this framework applies broadly to settings where individuals or firms face discrete threshold-based incentives, including welfare cliffs in transfer programs and regulatory size thresholds.

Second, we demonstrate that standard difference-in-differences methods yield biased estimates when policy thresholds induce units to relocate across cutoffs. To address this challenge, we develop the switcher difference-in-differences estimator, which can identify causal effects in situations where such relocation occurs. This methodology offers a tractable solution in settings where conventional approaches fail due to policy-induced mobility, with potential applications including, for example, welfare cliffs affecting labor supply, asset tests shaping wealth accumulation, and size-based regulations influencing firm growth.

Third, applying our framework to Finnish administrative data linking employers to employees and owners, we document two key findings. The payroll tax notch acted as a significant growth barrier, reducing the number of firms above the threshold by approximately 18 percent, with effects extending to firms eight times the threshold size. Following the notch's repeal, firm scale increased by roughly 10 percent across employment, capital stock, and value added, whereas standard difference-in-differences estimators would suggest much smaller or even negligible effects.

These results have important implications for policy design. Size-based thresholds are widespread in tax and regulatory systems, often justified as ways to reduce compliance burdens for small firms or target benefits efficiently. Our findings demonstrate that such policies can generate substantial distortions extending well beyond local effects around the cutoff. Policymakers should account for these broader distributional consequences when designing threshold-based policies, particularly if thresholds affect long-run growth trajectories. The theoretical and methodological tools we develop provide a framework for quantifying these effects across many policy contexts.

References

Abadie, A., J. Angrist, B. Frandsen, and J.-S. Pischke (2025): "Harvesting Differences-in-Differences and Event-Study Evidence," *NBER Working Paper No. 34550.*

Aghion, P., A. Bergeaud, and J. Van Reenen (2023): "The Impact of Regulation on Innovation," *American Economic Review*, 113, 2894–2936.

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

Benhabib, J. and A. Bisin (2018): "Skewed Wealth Distributions: Theory and Empirics," *Journal of economic literature*, 56, 1261–1291.

Benzarti, Y. and J. Harju (2021): "Using Payroll Tax Variation to Unpack the Black Box of Firm-Level Production," *Journal of the European Economic Association*, 19, 2737–2764.

Best, M., A. Brockmeyer, H. Kleven, J. Spinnewijn, and M. Waseem (2015): "Production vs Revenue Efficiency With Limited Tax Capacity: Theory and Evidence From Pakistan," *Journal of Political Economy*, 123, 1311–1355.

Cottet, S. (2023): "Payroll Tax Reductions for Minimum Wage Workers: Relative Labor Cost or Cash Windfall Effects?" *Working Paper.*

de Chaisemartin, C. and X. D'Haultfœuille (2023): "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey," *The econometrics journal*, 26, C1–C30.

DIXIT, A. K. (1993): *The Art of Smooth Pasting*, Harwood Academic Publishers.

GABAIX, X. (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1, 255–294.

——— (2016): "Power Laws in Economics: An Introduction," *Journal of Economic Perspectives*, 30, 185–206.

GARBINTI, B., J. GOUPILLE-LEBRET, M. MUÑOZ, S. STANTCHEVA, AND G. ZUCMAN (2023): "Tax Design, Information, and Elasticities: Evidence From the French Wealth Tax," .

GARICANO, L., C. LELARGE, AND J. VAN REENEN (2016): "Firm Size Distortions and the Productivity Distribution: Evidence from France," *American Economic Review*, 106, 3439–79.

GOURIO, F. AND N. ROYS (2014): "Size-dependent regulations, firm size distribution, and reallocation," *Quantitative Economics*, 5, 377–416.

GRUBER, J. (1997): "The Incidence of Payroll Taxation: Evidence from Chile," *Journal of Labor Economics*, 15.

GRUBER, J. AND E. SAEZ (2002): "The elasticity of taxable income: evidence and implications," *Journal of Public Economics*, 84, 1–32.

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.

HE (1972): "Valtiovarainvaliokunnan mietintö n:o 46 hallituksen esityksen johdosta laiksi leimaverolain muuttamisesta (available in Finnish only)," *Hallituksen esitys no. 172 / 1972*.

——— (2009): "Hallituksen esitys Eduskunnalle työnantajan kansaneläkemaksun poistamista koskevaksi lainsäädännöksi (available in Finnish only)," *Hallituksen esitys no. 147 / 2009*.

KLEVEN, H. J. (2016): "Bunching," *Annual Review of Economics*, 8, 435–64.

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.

OECD (2024): "Consumption Tax Trends 2024: VAT/GST and Excise, Core Design Features and Trends," *OECD Publishing, Paris.*

——— (2025): "Unleashing SME Potential to Scale Up: Helping SMEs Scale Up, OECD Studies on SMEs and Entrepreneurship," *OECD Publishing, Paris.*

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

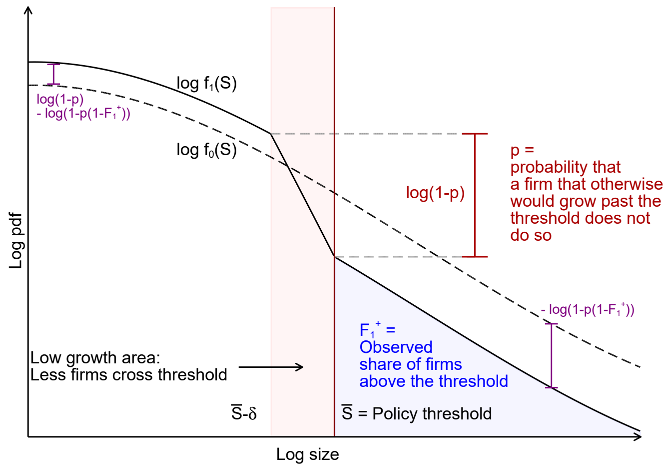
Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023): "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of econometrics*, 235, 2218–2244.

Saez, E., B. Schoefer, and D. Seim (2019): "Payroll taxes, firm behavior, and rent sharing: Evidence from a young workers' tax cut in Sweden," *American Economic Review*, 109, 1717–63.

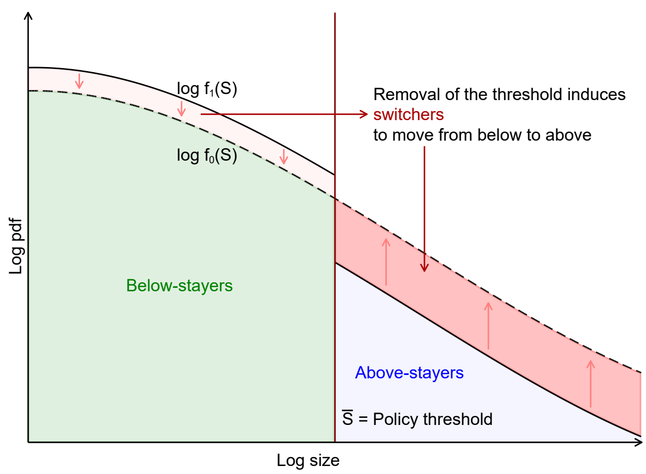
Stokey, N. L. (2009): *The Economics of Inaction: Stochastic Control Models with Fixed Costs*, Princeton University Press.

Figures

(a) Shifts in the firm distribution due to a growth barrier

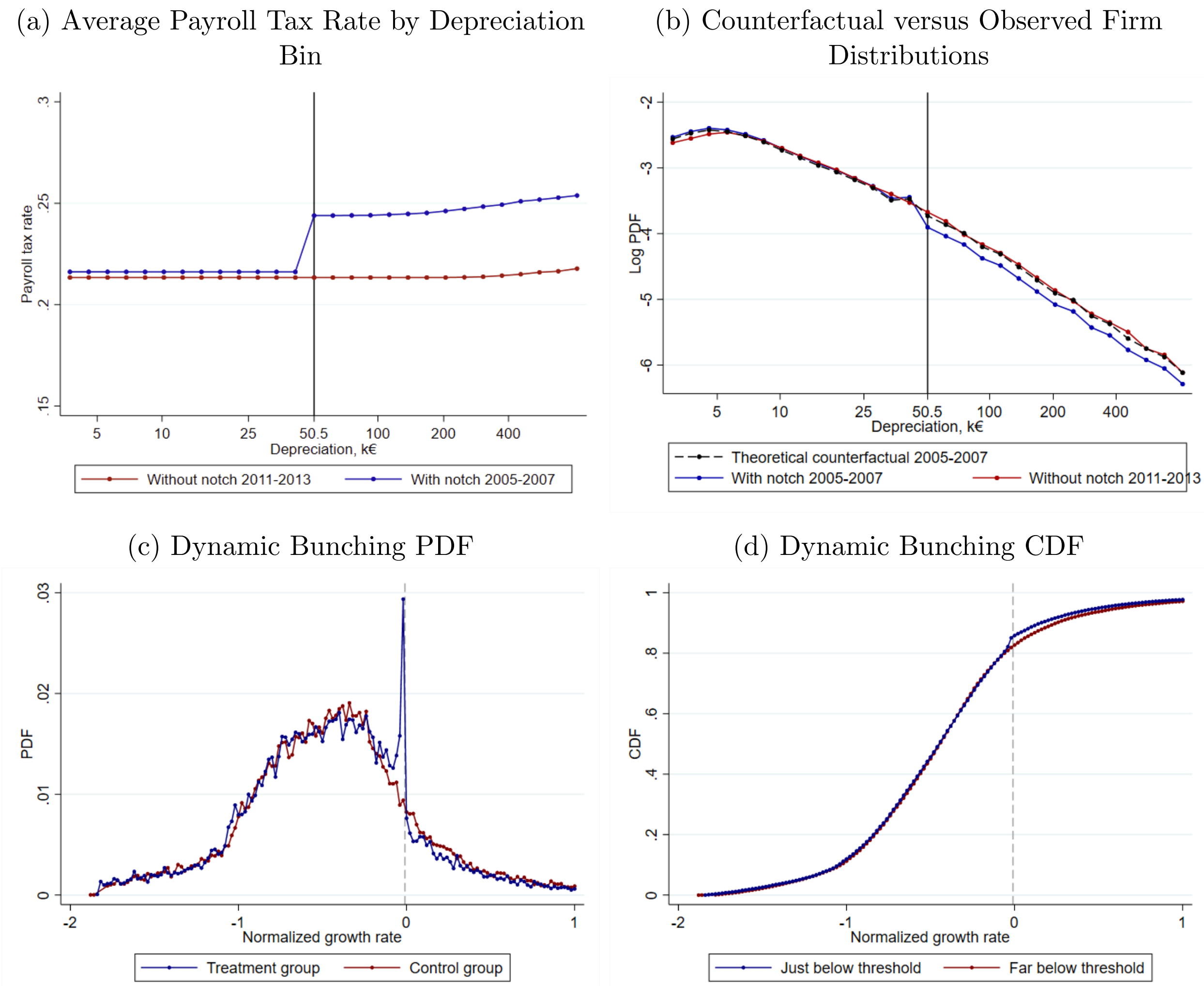


(b) Switchers and stayers



Notes: Figure 1a illustrates the shifts in the firm distribution due to a growth barrier in log-log space. We assume that the policy threshold at $\bar{S}$ creates a (small) low growth area $[\bar{S}-\delta, \bar{S}]$. p is the probability that a firm refrains from crossing the threshold due to the notch. F_1^+ is the share of firms above the threshold when the threshold is in place. The threshold shifts the distribution below the threshold up, and the distribution above the threshold down. These shifts are a function of p. This result holds (approximately) for a large family of dynamic systems of firm growth (see more details in Section 2).
Figure 1b shows how different groups of firms react to the removal of the notch. The solid black line shows the firm distribution when the threshold is in place. The dashed black line shows the distribution after the repeal of the notch. Removing the threshold induces switchers to move from below the notch (light red area) to above the notch (red area). Below-stayers (green area) and above-stayers (blue area) do not relocate, but only above-stayers are defined as treated as they are subject to the treatment. Hence, the treatment group consists of switchers and above-stayers. Standard difference-in-differences methods assume that there are only below- and above-stayers.

Figure 1: Shifts in the firm distribution and switchers



(a) Average Payroll Tax Rate by Depreciation Bin

(b) Counterfactual versus Observed Firm Distributions

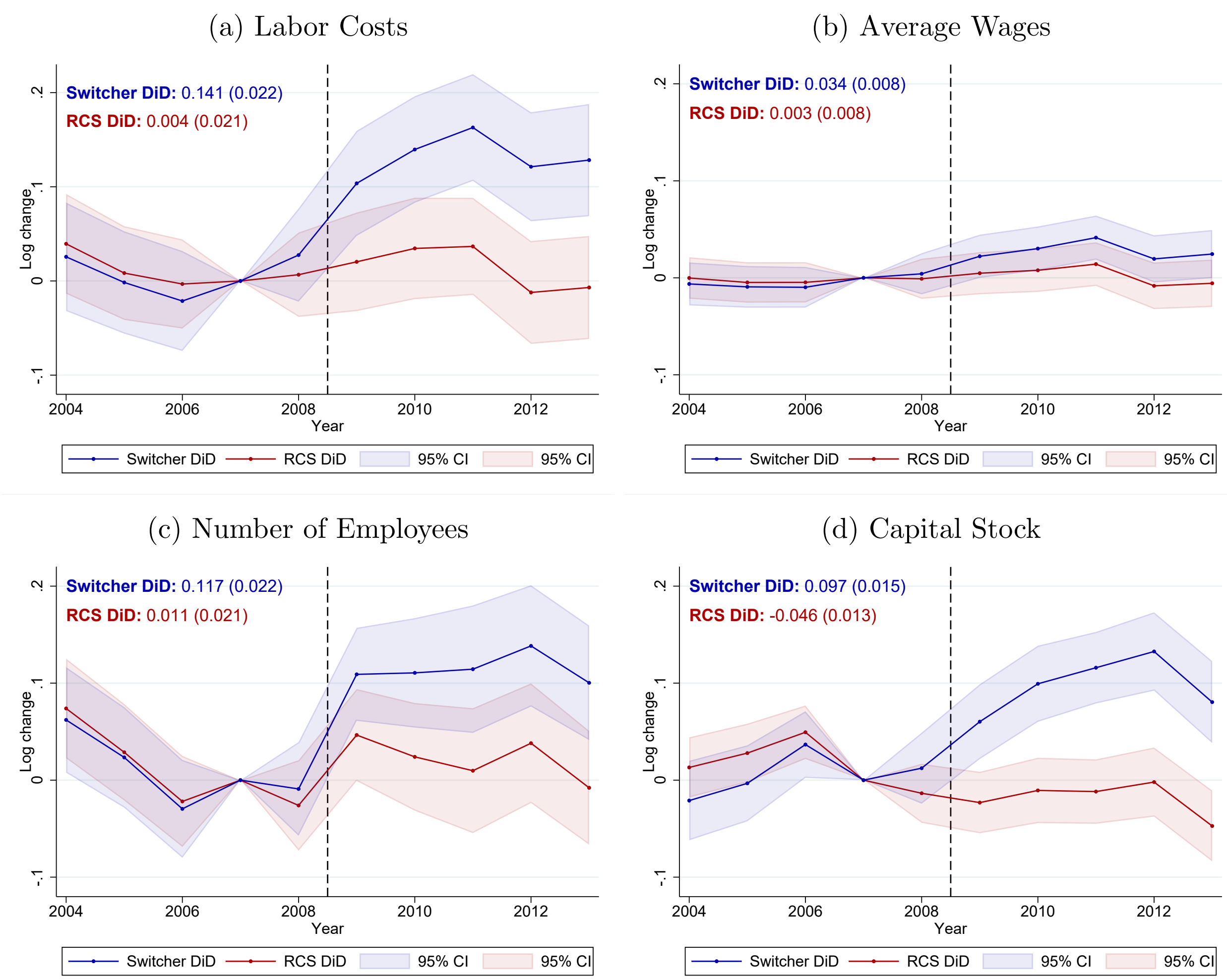
(c) Dynamic Bunching PDF

(d) Dynamic Bunching CDF

Notes: Panel 2a plots the variation of payroll tax rates in bins with (2005–2007) and without (2011–2013) the notch across the capital depreciation distribution. Panel 2b plots the log PDFs of firms in bins with (2005–2007) and without (2011–2013) the notch, and compares these to the theoretical counterfactual calculated based solely on pre-period (2005–2007) data. The width of the bins in panels 2a and 2b is calculated such that each bin captures 20 log points of capital depreciation up and down starting from the threshold. Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1)

Panels 2c and 2d show the normalized growth rate distributions of the dynamic bunching treated (depreciation between €27,715 and €50,499) and control (depreciation between €15,210 and €27,714) groups used in estimation of the probability that firms refrained from crossing the notch (p). The distributions of the control group have been scaled by a scaling parameter of 0.535 to provide a counterfactual distribution for the treated group. The gray vertical line shows the position of the threshold. Bins have a width of 2%. Treatment group bunching at the threshold is clearly visible in both the PDF and the CDF indicating that firms decide not to grow past the threshold in the treatment group. This implies lower growth rates for the firms in the treated group compared to the situation without the notch. The estimate for the probability that a firm refrains from crossing the notch is $p = 0.182$ with a bootstrapped standard error of 0.001.

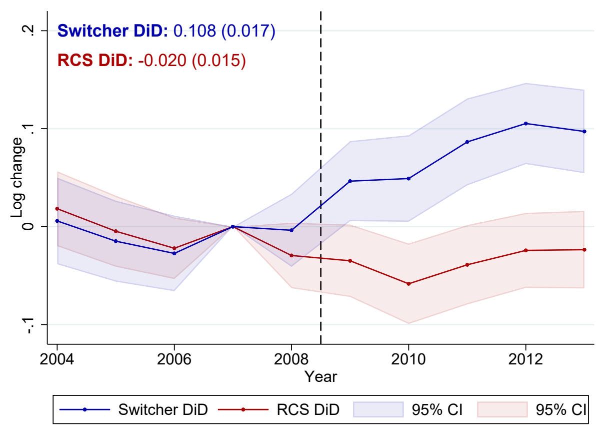
Figure 2: Payroll Tax Variation, Firm Capital Depreciation Distributions, and Dynamic Bunching Distributions



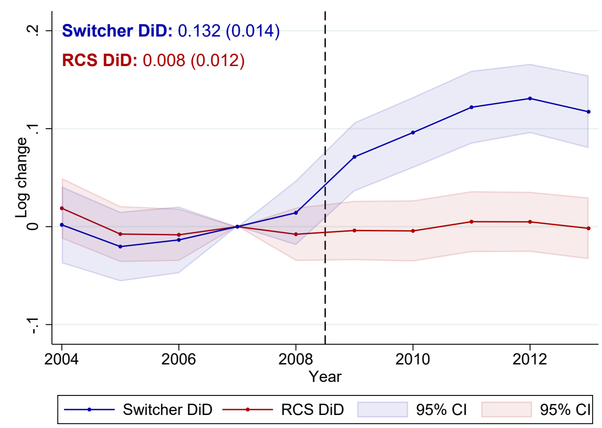
Notes: Panel 3a shows the event studies for switcher difference-in-differences (in blue) and repeated cross-sections difference-in-differences (in red) for log labor costs (see Section 3.5 for further details on the methodology). The panel also shows the switcher DiD and repeated cross-sections DiD point estimates. We compare firms 100 log points above the notch (annual capital depreciation €50,500 – €137,273) and 100 log points below (annual capital depreciation €18,578 – €50,499). Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1). For the point estimates, we compare years 2005–2007 and 2011–2013. Standard errors and confidence intervals are bootstrapped (500 resamples) by clustering at the firm level.
Panels 3b, 3c, and 3d show similar estimates for log averages wages, log number of employees, and log capital stock.

Figure 3: Event-Studies: Labor Costs, Average Wages, Number of Employees, and Capital Stock

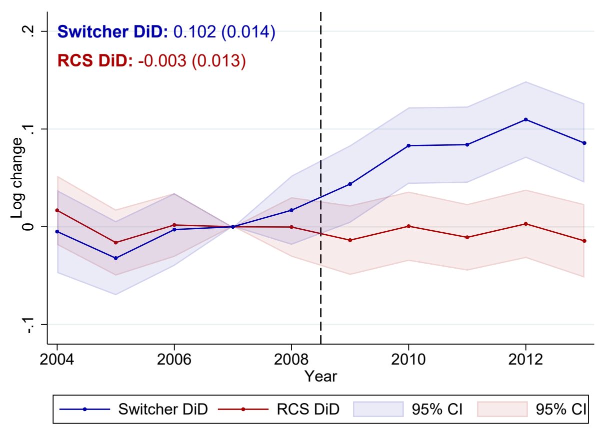
(a) Revenue



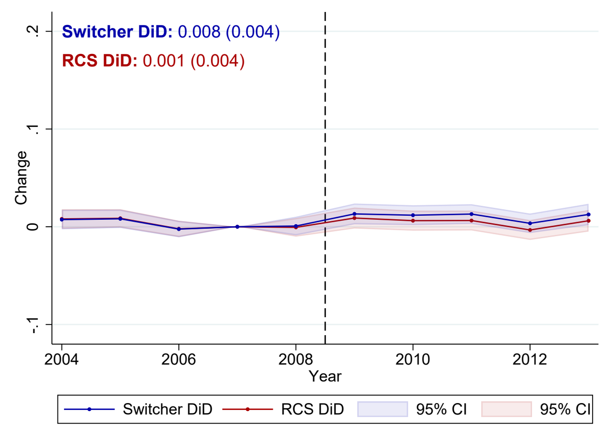
(b) Value Added



(c) Profits



(d) Labor Share



Notes: Panel 4a shows the event studies for switcher difference-in-differences (in blue) and repeated cross-sections difference-in-differences (in red) for log revenue (see Section 3.5 for further details on the methodology). The panel also shows the switcher DiD and repeated cross-sections DiD point estimates. We compare firms 100 log points above the notch (annual capital depreciation €50,500 – €137,273) and 100 log points below (annual capital depreciation €18,578 – €50,499). Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1). For the point estimates, we compare years 2005–2007 and 2011–2013. Standard errors and confidence intervals are bootstrapped (500 resamples) by clustering at the firm level. Panels 4b, 4c, and 4d show similar estimates for log value added, log profits, and labor share.

Figure 4: Event-Studies: Revenue, Value-Added, Profits, and Labor Share

Tables

	All	Baseline sample	Last 20 log points below the threshold
Capital Depreciation in taxation	62,772	44,768	44,614
Capital Depreciation in accounting	45,922	45,195	46,207
Labor Costs	230,818	189,051	208,647
Number of Employees	7.69	5.18	6.71
Routine	3.30	2.93	2.57
Non-Routine	4.16	3.78	3.96
Graduate	1.24	0.52	0.50
Non-Graduate	6.45	5.81	6.22
Male	4.81	4.49	4.79
Female	2.81	2.37	1.86
Revenue	1,336,311	807,312	875,322
Value added	323,297	264,360	288,929
Capital stock	896,150	353,366	329,762
Total Investment	94,165	81,538	87,824
Investment in fixed assets	49,351	61,080	69,368
Investment in buildings	13,272	10,191	10,130
Investment in R&D	28,356	5,114	3,580
Rental Costs	30,862	16,582	16,927
Services bought outside	97,117	39,060	41,041
Dividends	98,702	45,179	23,248
Number of firms	702,103	65,807	7,811

Notes: Table 1 presents the mean value of each variable used in our analysis for three different samples: All firms, our baseline sample, and last bin on the left of the threshold in our pre-reform period (i.e. years 2005–2007). The baseline sample includes firms that had capital depreciation within 100 log points from the threshold (between €18,578 and €137,273), and were affected by the reform. The last column has firms within 20 log points below the threshold (between €41,348 and €50,499) that were affected by the reform (see Section 4.1 for more details on which firms were affected).

Table 1: Descriptive statistics: Years 2005–2007

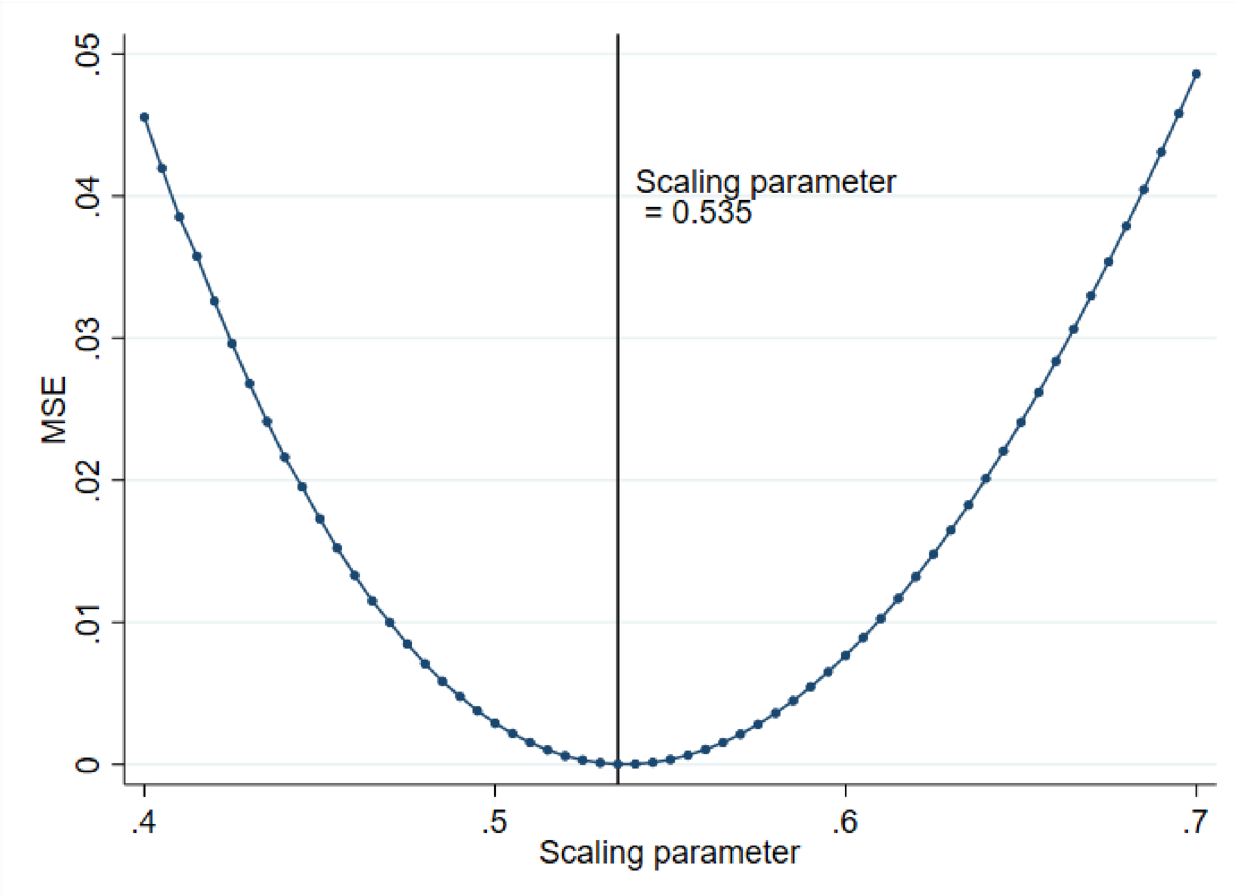
	(1) Labor Costs	(2) Average Wages	(3) Number of Employees	(4) Capital Stock	(5) Revenue	(6) Value Added	(7) Profits	(8) Labor Share
DiD estimates								
Switcher DiD	0.141	0.034	0.117	0.097	0.108	0.132	0.102	0.008
	(0.022)	(0.008)	(0.022)	(0.015)	(0.017)	(0.012)	(0.014)	(0.004)
RCS DiD	0.004	0.003	0.011	−0.046	−0.020	0.008	−0.003	0.001
	(0.021)	(0.008)	(0.021)	(0.013)	(0.015)	(0.010)	(0.013)	(0.004)
Switcher term	0.137	0.031	0.105	0.143	0.128	0.124	0.105	0.006
	(0.010)	(0.002)	(0.007)	(0.009)	(0.009)	(0.009)	(0.008)	(0.001)
Switcher share	0.126	0.131	0.129	0.135	0.130	0.131	0.128	0.133
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
FE DiD	−0.173	−0.043	−0.114	−0.221	−0.154	−0.169	−0.185	0.011
	(0.015)	(0.007)	(0.012)	(0.011)	(0.011)	(0.011)	(0.014)	(0.003)
Observations	132,935	114,389	133,076	131,394	136,331	130,221	119,191	137,026
Theory-based estimates								
Theory based switcher term	0.188	0.039	0.142	0.183	0.172	0.165	0.143	0.009
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.000)
Theory based switcher share	0.173	0.167	0.175	0.173	0.175	0.175	0.175	0.175
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	60,852	56,676	64,829	63,332	65,807	63,451	58,855	65,807

Notes: Difference-in-differences estimates, the components of switcher difference-in-differences for different outcomes, and theory based estimates of the switcher term. We compare firms 100 log points above the notch (annual capital depreciation €50,500 – €137,273) and 100 log points below (annual capital depreciation €18,578 – €50,499) before (2005–2007) and after (2011–2013) the removal of the notch. Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1). Standard errors in parentheses have been bootstrapped (500 resamples) and they are clustered at the firm level.

Table 2: Difference-in-differences and theory-based point estimates

A Appendix – Additional Figures and Tables

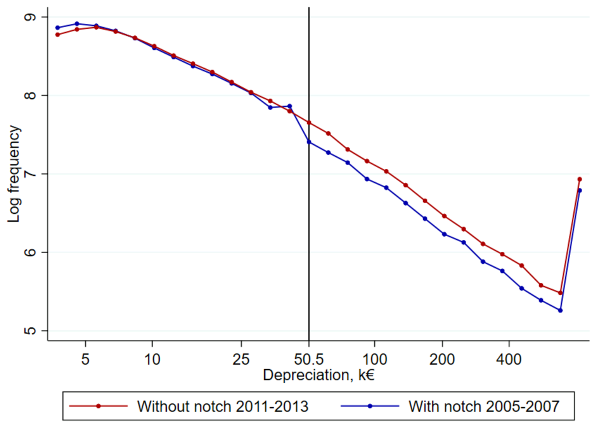
Figures



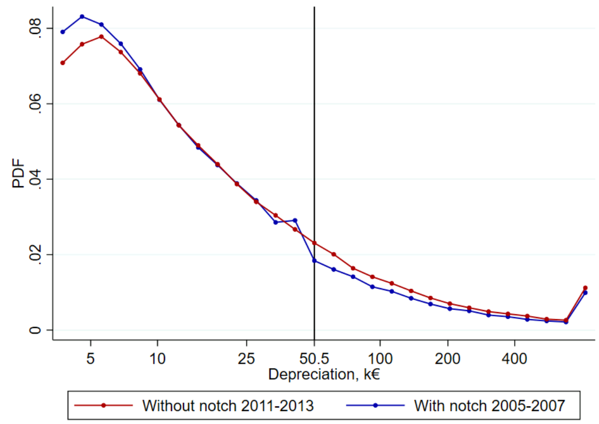
Notes: This figure shows the mean squared errors (MSE) of matching the cumulative growth rate distributions of the dynamic bunching control group (depreciation between €15,210 and €27,714) to the dynamic bunching treated group (depreciation between €22,715 and €50,499) in our dynamic bunching setup. The distributions are matched for growth rate bins below -4% below the threshold (true threshold for treated group, placebo threshold for control group). The MSEs are shown for different scaling parameters, spaced between 0.005 units. This data-driven approach to choosing the scaling parameter suggests using a scaling parameter of around 0.535.

Figure A1: Choosing scaling parameter for dynamic bunching

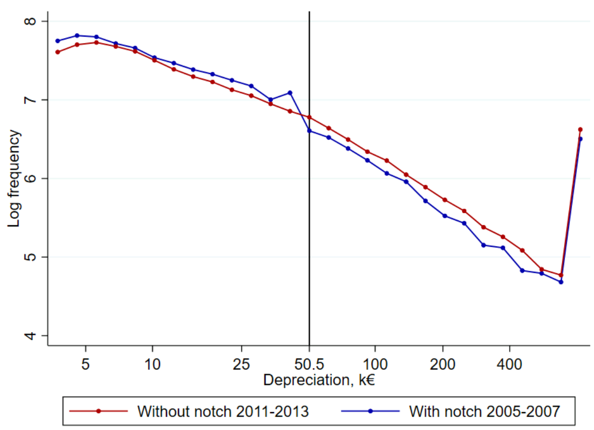
(a) Log distribution of firms



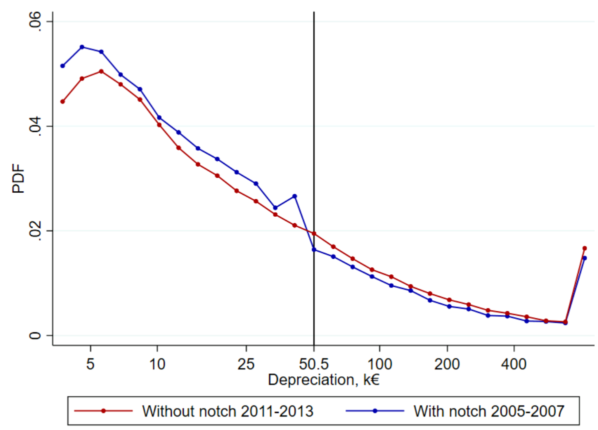
(b) Density of firms



(c) Log distribution of firms, balanced panel



(d) Density of firms, balanced panel

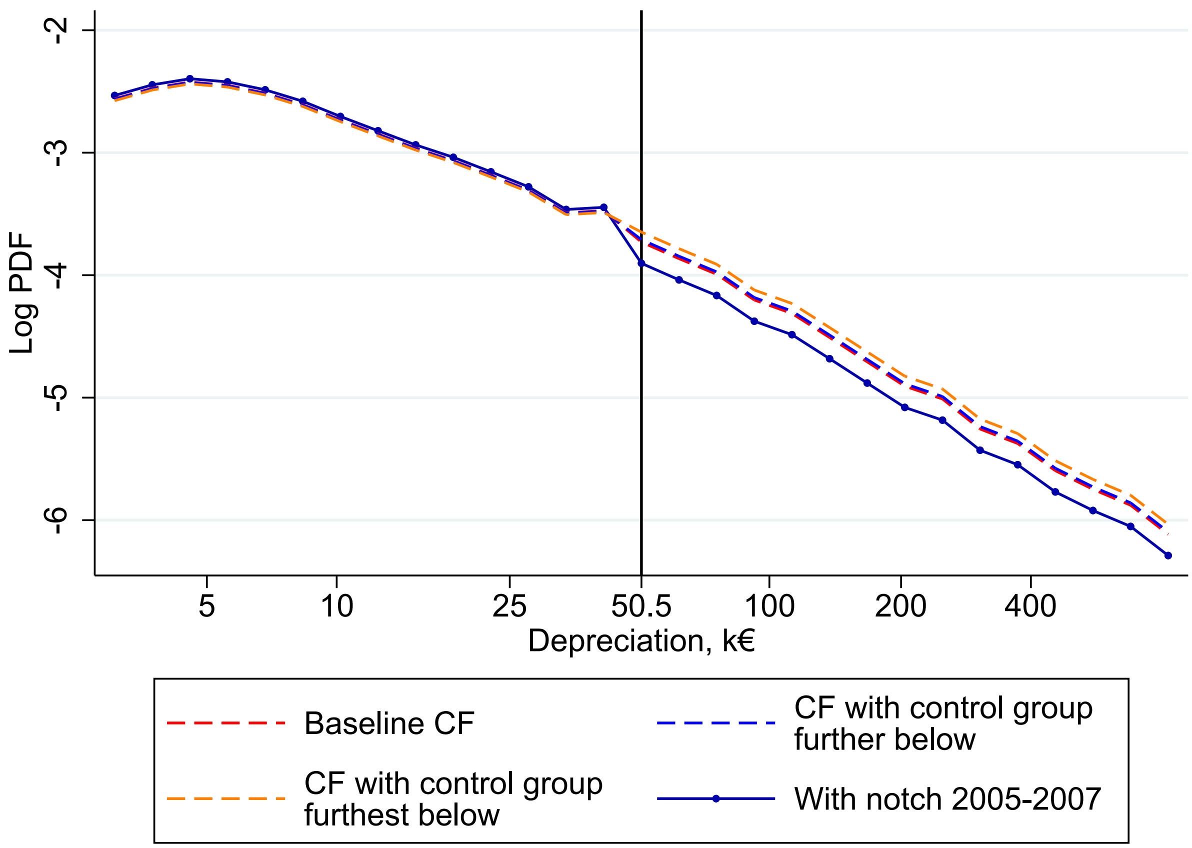


Notes: Figure A2a plots the log frequency of firms in bins with (2005–2007) and without (2011–2013) the notch across the capital depreciation distribution. The width of the bins is calculated such that each bin captures 20 log points of capital depreciation up and down starting from the threshold. Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1). Figure A2b plots the density of firms in bins with (2005–2007) and without (2011–2013) the notch across the capital depreciation distribution. Figures A2c and A2d show the upper-panel distributions for a balanced panel of firms.

Figure A2: Firm distributions: Pre- and Post-2010, Full Sample and Balanced Panel

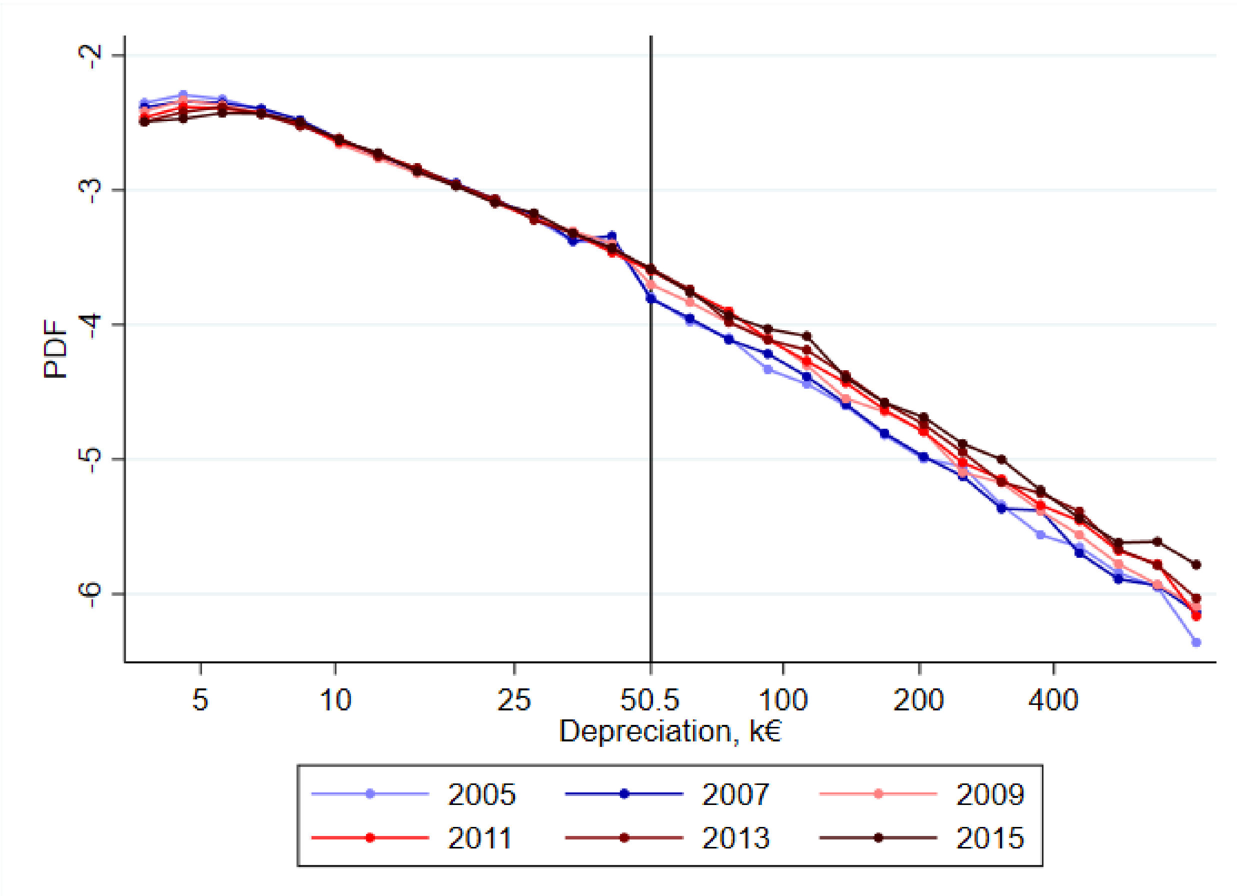
Notes: Figure A3 plots the average likelihood of firms staying in the same bin (Figure A3a), moving to some higher bin (Figure A3b), moving to some bin lower than where they were (Figure A3c), or exiting the sample (Figure A3d) between years 2002–2004 and 2005–2007 (with notch), and years 2011–2013 and 2014–2016 (without notch). The width of the bins is calculated such that each bin captures 20 log points of capital depreciation up and down starting from the threshold.

Figure A3: Firm movement in the distribution.



Notes: This figure plots the log PDFs of firms in bins with (2005–2007) the notch (solid blue line), and compares these to the theoretical counterfactual distributions calculated based solely on pre-period (2005–2007) data, but with different control groups. The baseline counterfactual (dashed red) uses as its dynamic bunching control group firms between 60 and 120 log points below the threshold (capital depreciation between €15,210 and €27,714). The counterfactual with a control further below (dashed blue) uses as its counterfactual firms between 120 and 180 log points below the threshold (capital depreciation between €8,347 and €15,210). The counterfactual with a control furthest below (dashed orange) uses as its counterfactual firms between 180 and 240 log points below the threshold (capital depreciation between €4,581 and €8,347). The figure indicates that the theoretical counterfactual is not sensitive to the choice of the dynamic bunching control group. The width of the bins in the figure is calculated such that each bin captures 20 log points of capital depreciation up and down starting from the threshold. Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1)

Figure A4: Counterfactual Distributions with Different Control groups



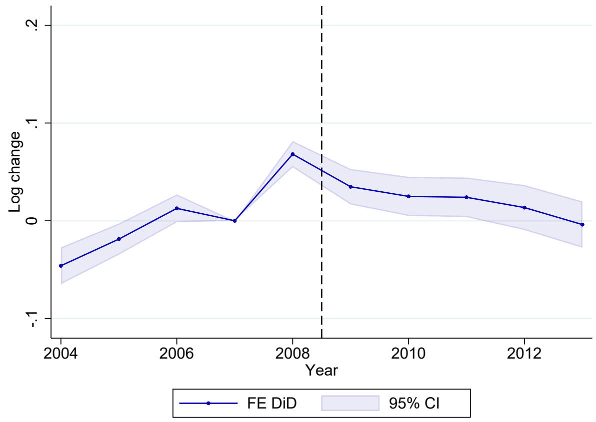
Notes: This figure plots the log PDFs of the firm distribution in bins every two years both with (blue lines, i.e. 2005 and 2007) and without (red lines, i.e. 2009–) the notch across the capital depreciation distribution. The width of the bins is calculated such that each bin captures 20 log points of capital depreciation up and down starting from the threshold. Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1). The figure indicates that it took several years for the distribution to "fill-up" on the right after the repeal of the notch.

Figure A5: Evolution of the Firm Distribution over Time

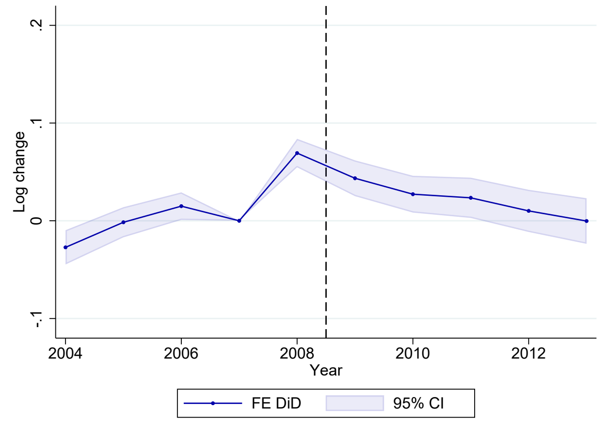
Notes: Panel A6a shows the event studies for firm fixed effects difference-in-differences for log labor costs. Treatment assignment is based on whether a given firm was above (treated) or below (control) the threshold in 2007. We compare firms 100 log points above the notch (annual capital depreciation €50,500 – €137,273) and 100 log points below (annual capital depreciation €18,578 – €50,499). Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1). Confidence intervals are bootstrapped (500 resamples) by clustering at the firm level. Panels A6b, A6c, and A6d show similar estimates for log averages wages, log number of employees, and log capital stock.

Figure A6: Event-Studies with Firm FE's: Labor Costs, Average Wages, Number of Employees, and Capital Stock

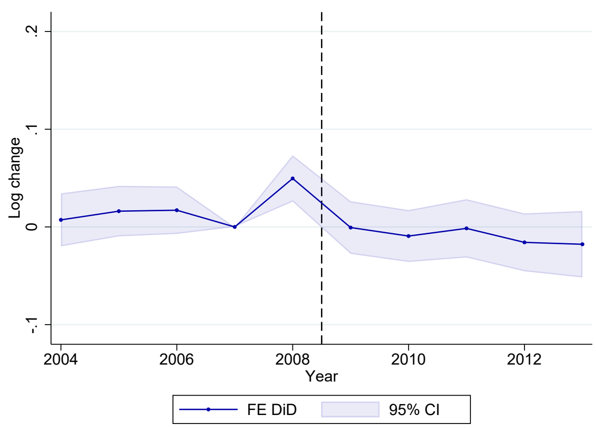
(a) Revenue



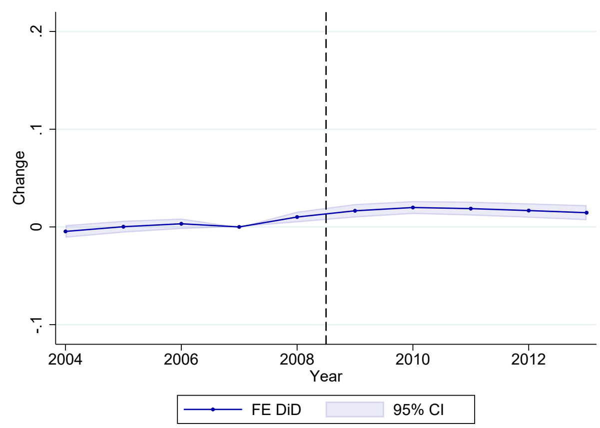
(b) Value Added



(c) Profits



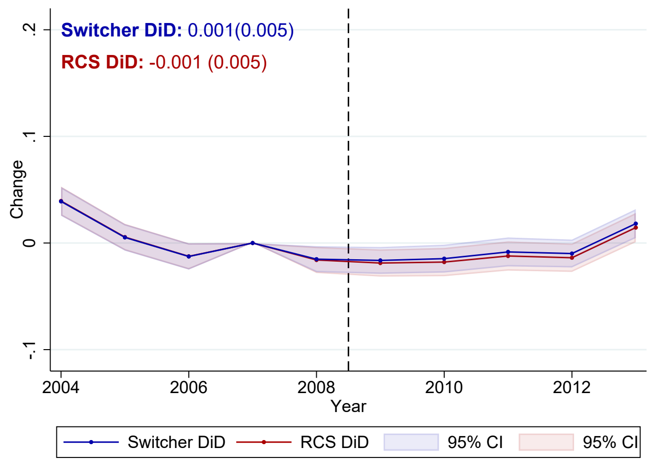
(d) Labor Share



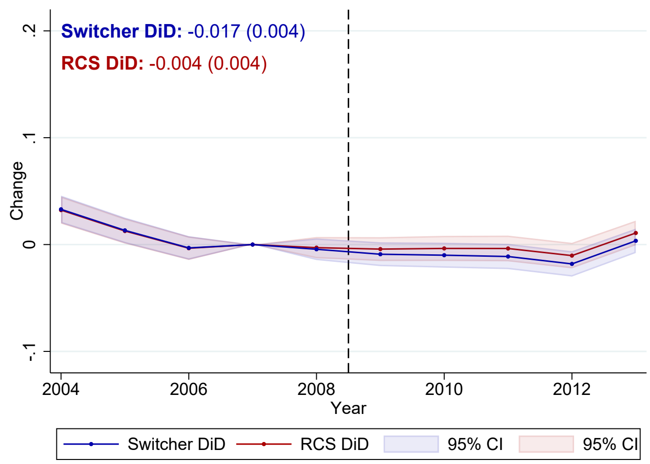
Notes: Panel A7a shows the event studies for firm fixed effects difference-in-differences for log revenue. Treatment assignment is based on whether a given firm was above (treated) or below (control) the threshold in 2007. We compare firms 100 log points above the notch (annual capital depreciation €50,500 – €137,273) and 100 log points below (annual capital depreciation €18,578 – €50,499). Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1). Confidence intervals are bootstrapped (500 resamples) by clustering at the firm level. Panels A7b, A7c, and A7d show similar estimates for log value added, log profits, and labor share.

Figure A7: Event-Studies with Firm FE's: Revenue, Value-Added, Profits, and Labor Share

(a) Share of Workers without Completed Tertiary Education

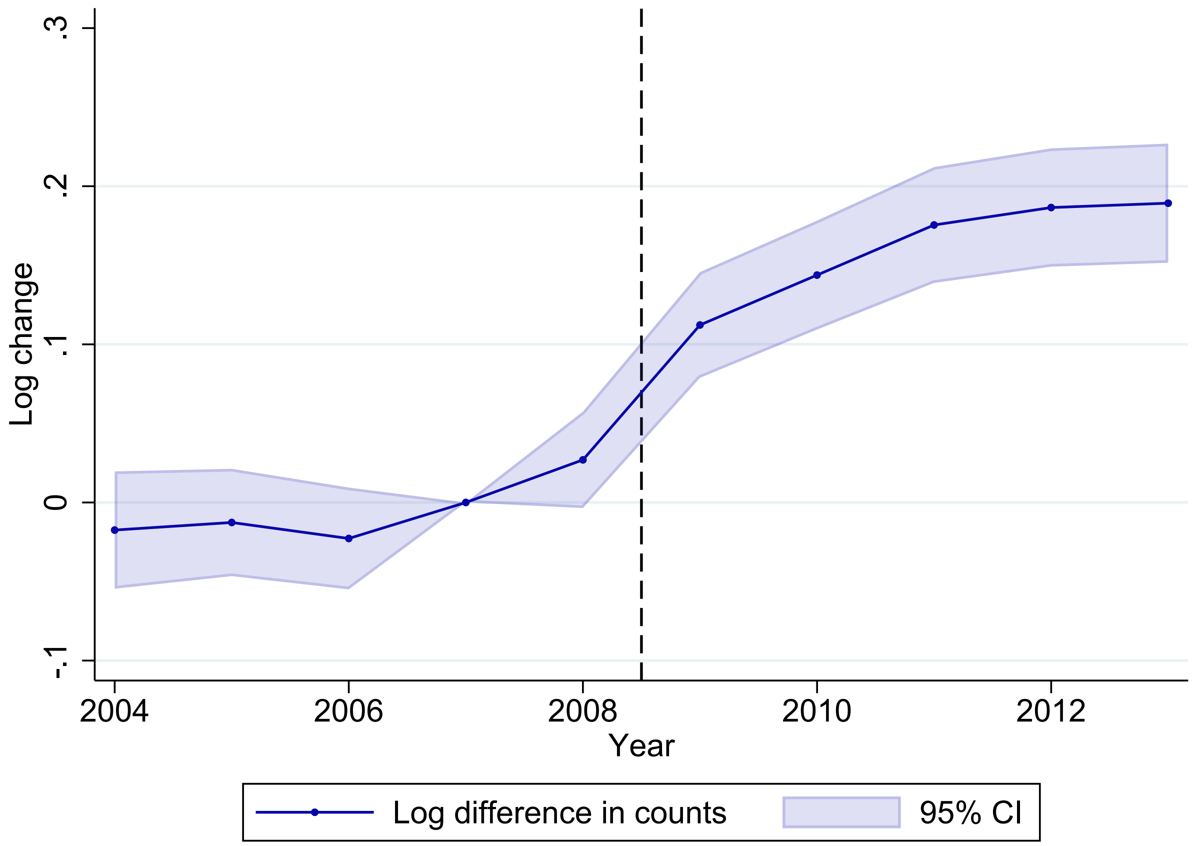


(b) Share of Routine Workers



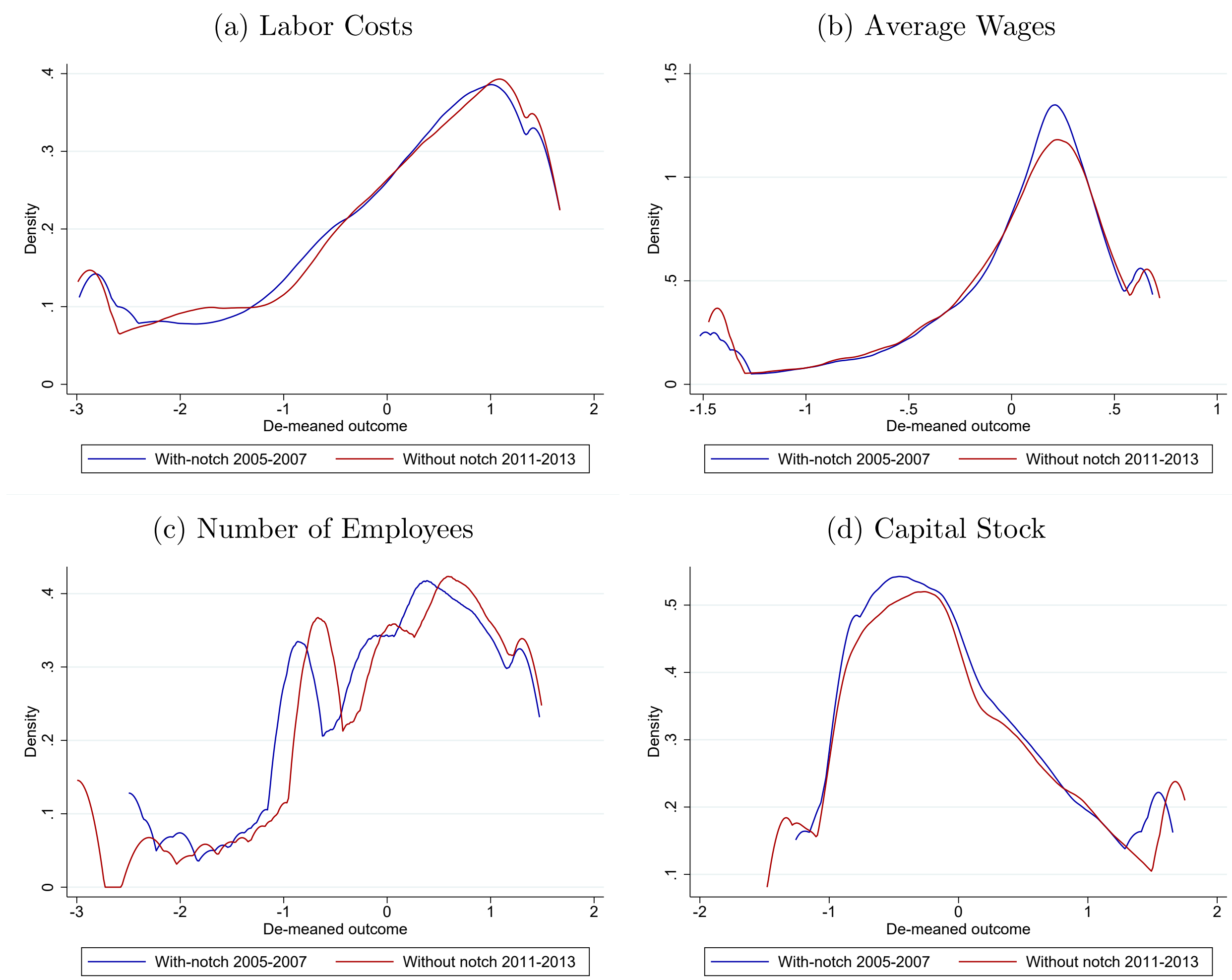
Notes: Panel A8a shows the event studies for switcher difference-in-differences (in blue) and repeated cross-sections difference-in-differences (in red) for the share of workers without a completed tertiary education (see Section 3.5 for further details on the methodology). The panel also shows the switcher DiD and repeated cross-sections DiD point estimates. We compare firms 100 log points above the notch (annual capital depreciation €50,500 – €137,273) and 100 log points below (annual capital depreciation €18,578 – €50,499). Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1). For the point estimates, we compare years 2005–2007 and 2011–2013. Standard errors and confidence intervals are bootstrapped (500 resamples) by clustering at the firm level. Panel A8b shows similar estimates for share of routine workers.

Figure A8: Worker Composition within Firms



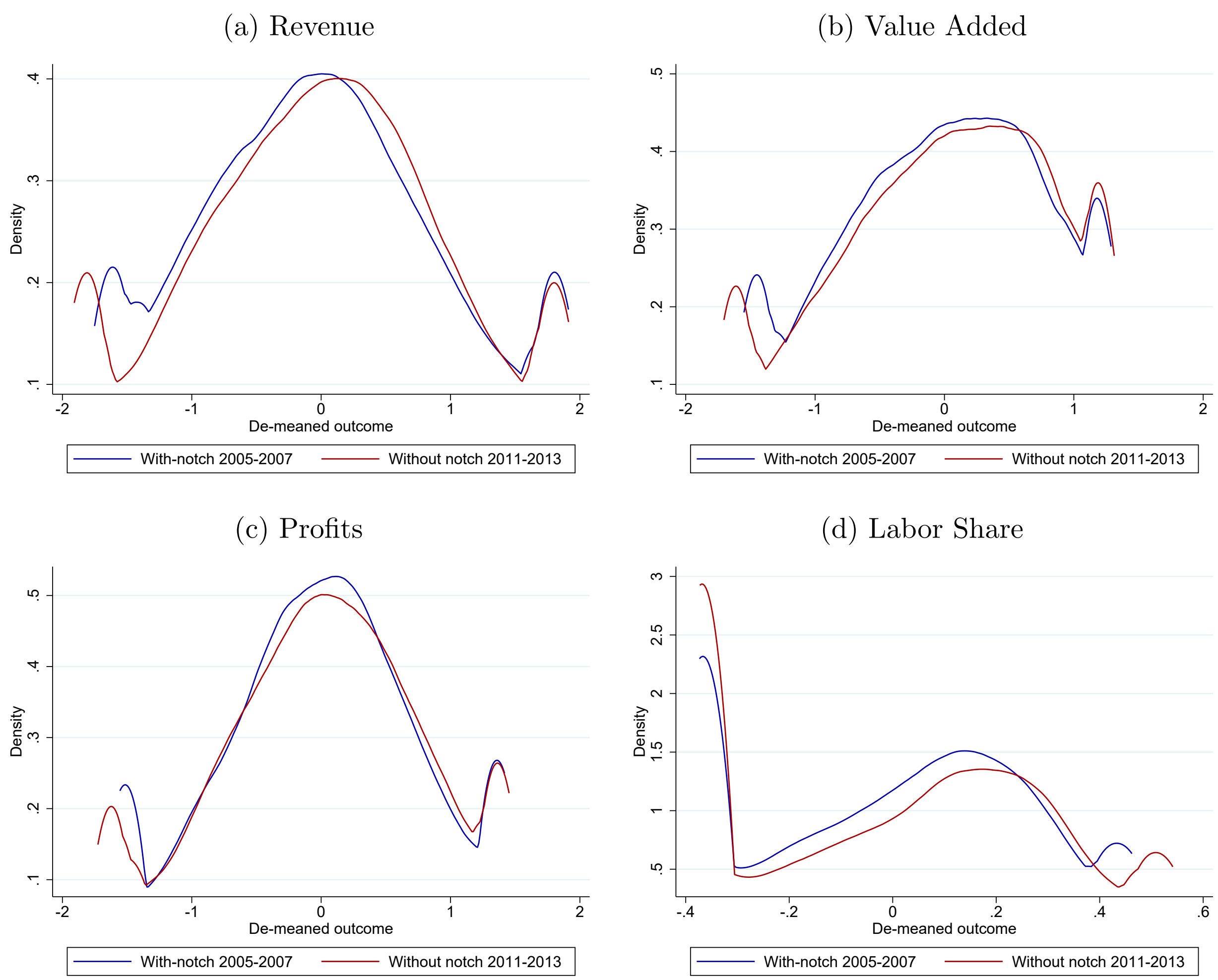
Notes: This figure plots the difference of log counts (i.e. number of firms) above versus below the payroll tax notch over time. The notch is at €50,500 of annual capital depreciation. Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1). Moreover, only firms with capital depreciation between around €18,600 and €137,300 are included, as in our main difference-in-difference specifications. This figure shows differences in log counts for labor costs, but the log counts for other outcomes look very similar, as the only difference is whether a given outcome has been observed for a given firm.

Figure A9: Evolution of log counts above versus below



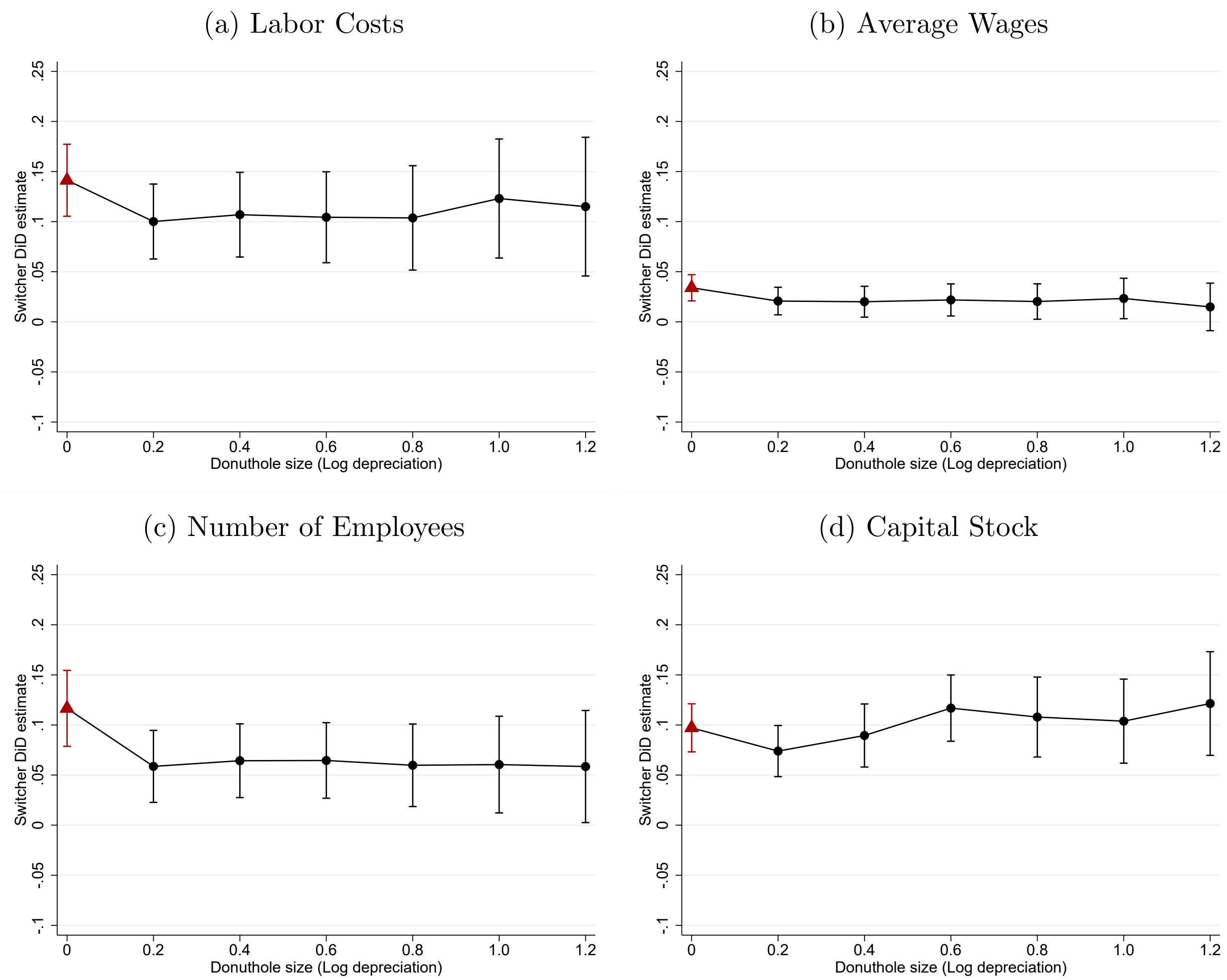
Notes: Panel A10a shows the kernel density estimates of the de-meaned log labor cost distributions below the capital depreciation threshold before (2005–2007) and after (2011–2013) the repeal of the notch. We include firms 100 log points below the notch. Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1). Panels A10b, A10c, and A10d show similar kernel density estimates for log averages wages, log number of employees, and log capital stock.

Figure A10: Change in De-meaned Outcome Distributions: Labor Costs, Average Wages, Number of Employees, and Capital Stock



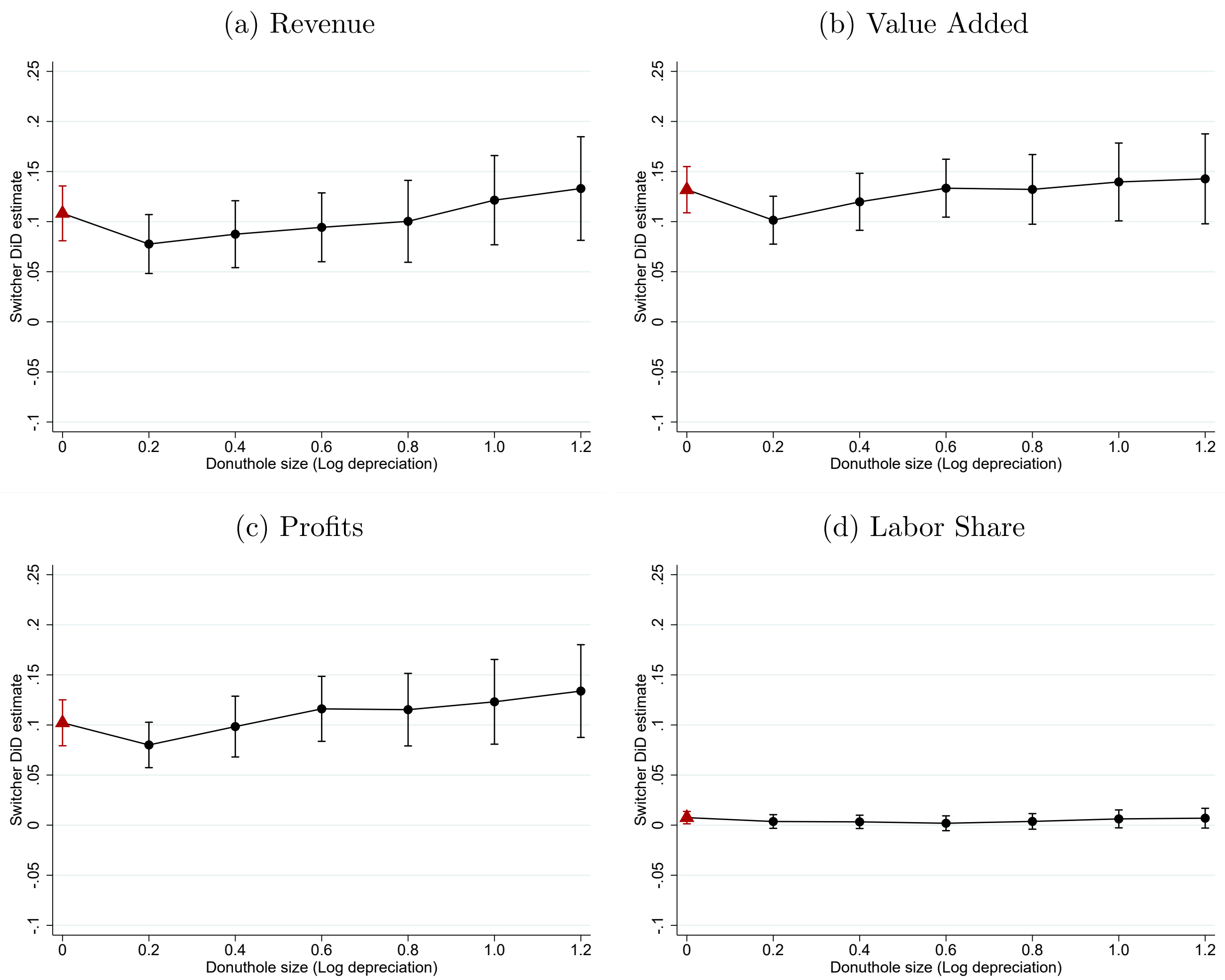
Notes: Panel A11a shows the kernel density estimates of the de-meaned log revenue distributions below the capital depreciation threshold before (2005–2007) and after (2011–2013) the repeal of the notch. We include firms 100 log points below the notch. Only firms that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to them (see more details on the institutional setup in Section 4.1). Panels A11b, A11c, and A11d show similar kernel density estimates for log value added, log profits, and labor share.

Figure A11: Event-Studies: Revenue, Value-Added, Profits, and Labor Share



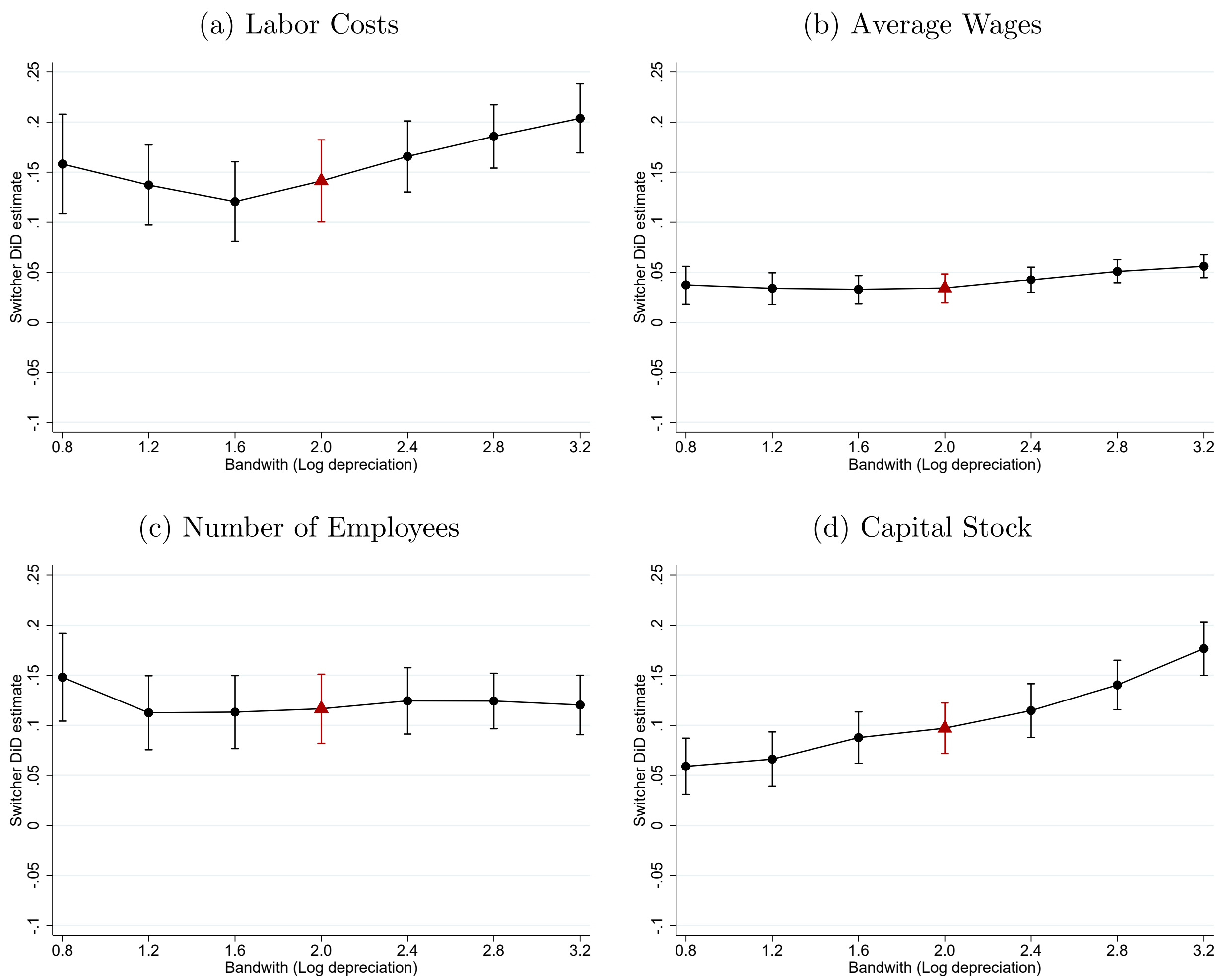
Notes: The figure shows how the switcher difference-in-differences estimates change when a donut hole is introduced near the threshold. The red triangle shows our baseline point estimate without a donut hole, while the black dots show estimates with otherwise the same specification, but with observations within a band near the threshold left out (donut hole).

Figure A12: Impact of Donut Hole Width on Switcher-DiD Estimates for Labor Costs, Average Wages, Number of Employees and Capital Stock



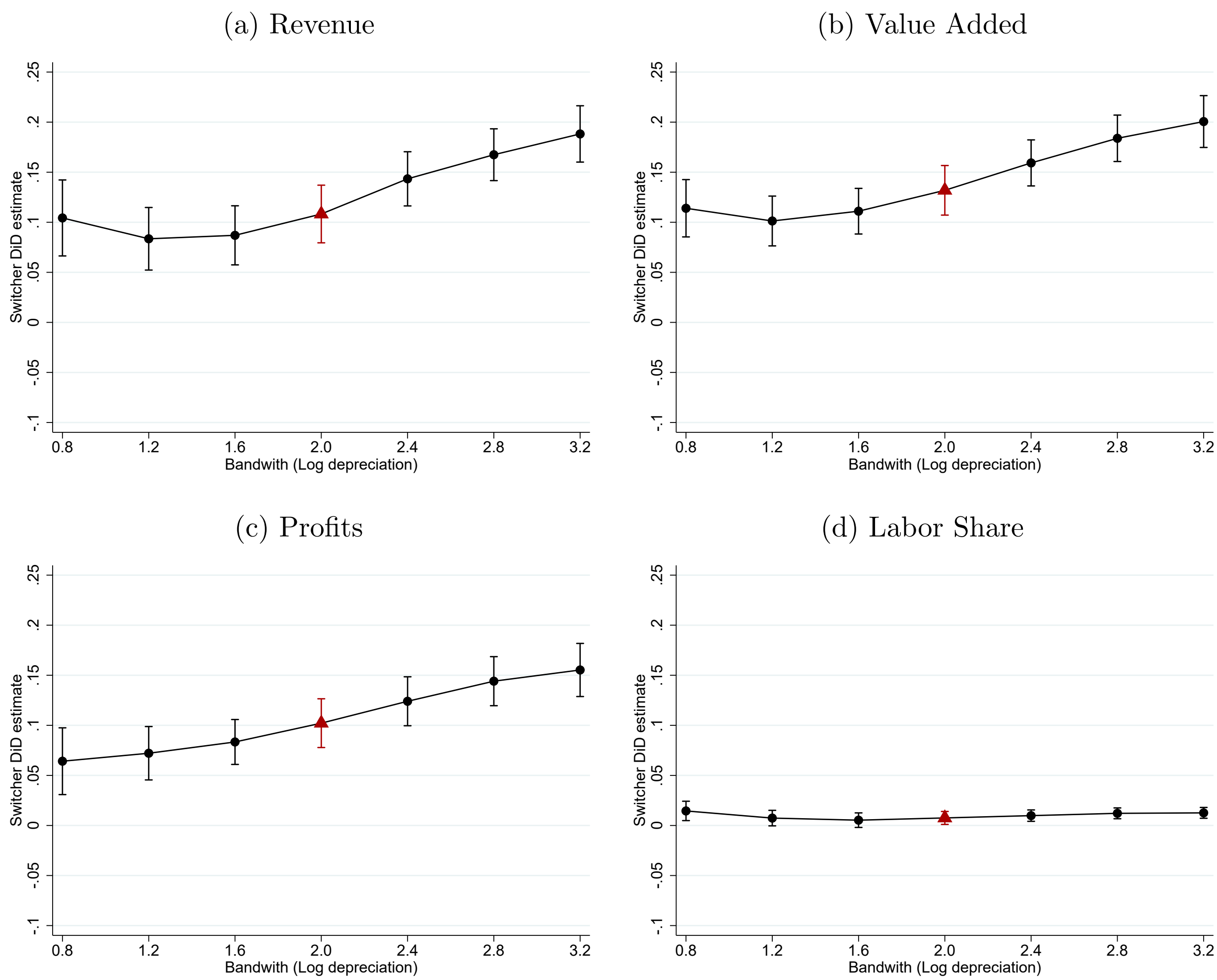
Notes: The figure shows how the switcher difference-in-differences estimates change when a donut hole is introduced near the threshold. The red triangle shows our baseline point estimate without a donut hole, while the black dots show estimates with otherwise the same specification, but with observations within a band near the threshold left out (donut hole).

Figure A13: Impact of Donut Hole Width on Switcher-DiD Estimates for Revenue, Value Added, Profits, and Labor Share



Notes: The figure shows how the switcher difference-in-differences estimates change when the bandwidth is changed, i.e. when observations further away from the threshold are included in the estimation sample. The red triangle shows our baseline point estimate, while the black dots show estimates with otherwise the same specification, but with different bandwidths.

Figure A14: Impact of Bandwidth on Switcher-DiD Estimates for Labor Costs, Average Wages, Number of Employees and Capital Stock



Notes: The figure shows how the switcher difference-in-differences estimates change when the bandwidth is changed, i.e. when observations further away from the threshold are included in the estimation sample. The red triangle shows our baseline point estimate, while the black dots show estimates with otherwise the same specification, but with different bandwidths.

Figure A15: Impact of Bandwidth on Switcher-DiD Estimates for Revenue, Value Added, Profits, and Labor Share

Tables

Definition for firm categories					
I	D $<$ 50,500	**or**	D $\geq$ 50,500	and	D $<$ 0.1 * labor costs
II	D $\geq$ 50,500	and	D $\geq$ 0.1 * labor costs	and	D $<$ 0.3 * labor costs
III	D $\geq$ 50,500	and	D $\geq$ 0.3 * labor costs		

Notes: D refers to tax-deductible capital depreciation and labor costs refer to all salaries.

Table A1: Firm categories for payroll tax rates

	Health and pension				Unemployment					
	Firm categories*			Accident	Firm categories**		Group life	Employers	Total	Total
Year	I	II	III	insurance***	I	II	insurance***	pension***	lowest	highest
1996	4.000	5.600	6.500	1.2	1.00	4.00	0.100	16.80	23.100	28.600
1997	4.000	5.600	6.500	1.4	1.00	4.00	0.090	16.70	23.190	28.690
1998	4.000	5.600	6.500	1.4	0.90	3.90	0.080	16.80	23.180	28.680
1999	4.000	5.600	6.500	1.3	0.90	3.85	0.080	16.80	23.080	28.530
2000	4.000	5.600	6.500	1.2	0.90	3.45	0.090	16.80	22.990	28.040
7/2000	3.600	5.600	6.500	1.2	0.90	3.45	0.090	16.80	22.590	28.040
2001	3.600	5.600	6.500	1.2	0.80	3.10	0.095	16.60	22.295	27.495
2002	3.600	5.600	6.500	1.1	0.70	2.70	0.095	16.70	22.185	27.085
3/2002	2.950	5.150	6.050	1.1	0.70	2.70	0.095	16.70	21.535	26.635
2003	2.964	5.164	6.064	1.1	0.60	2.45	0.081	16.80	21.545	26.495
2004	2.964	5.164	6.064	1.1	0.60	2.50	0.080	16.80	21.544	26.544
2005	2.966	5.166	6.066	1.2	0.70	2.80	0.080	16.80	21.746	26.946
2006	2.958	5.158	6.058	1.1	0.75	2.95	0.080	16.70	21.588	26.888
2007	2.951	5.151	6.051	1.1	0.75	2.95	0.080	16.64	21.521	26.821
2008	2.771	4.971	5.871	1.0	0.70	2.90	0.080	16.80	21.351	26.651
2009	2.801	5.001	5.901	1.0	0.65	2.70	0.070	16.80	21.321	26.471
4/2009	2.000	4.201	5.101	1.0	0.65	2.70	0.070	16.80	20.520	25.601
2010	2.220	2.220	2.220	0.8	0.75	2.95	0.070	16.90	20.74	22.94
2011	2.210	2.210	2.210	1.0	0.80	3.20	0.070	17.10	21.18	23.58
2012	2.210	2.210	2.210	1.0	0.80	3.20	0.070	17.35	21.43	23.83
2013	2.040	2.040	2.040	0.9	0.80	3.15	0.070	17.35	21.16	23.51
2014	2.140	2.140	2.140	0.9	0.75	2.95	0.070	17.75	21.61	23.81
2015	2.080	2.080	2.080	0.9	0.80	3.15	0.070	18.00	21.85	24.89
2016	2.120	2.120	2.120	0.8	1.0	3.90	0.070	18.00	21.99	24.89
2017	1.080	1.080	1.080	0.8	0.8	3.30	0.070	17.95	20.70	23.20

Notes: * Refers to firm categories by labor costs and capital depreciation.
** Category I is for wages below a certain wage-sum threshold, e.g., €2,059,500 in year 2017, and Category II is for wages above the threshold. The threshold varies slightly over the years.
*** Represents the average values of these insurances.

Table A2: Social insurance percentages by firm categories, insurance types and years

B Appendix - Derivations for Section 2

Our aim is to derive an expression for the stationary firm distribution with and without the growth reduction in Equation (5).

No frictions In the no-friction case, the forward Kolmogorov Equation (or the Fokker-Planck Equation) is everywhere simply

$$\partial_t f(S_t) = \partial_S g(S_t) f(S_t) - \frac{1}{2}\partial_{SS}\left[\sigma^2(S_t) f(S_t)\right], \tag{20}$$

where $\partial_t f(S_t) = \partial f(S_t)/\partial t$, $\partial_S g(S_t) f(S_t) = \partial g(S_t)/\partial S_t$, and $\partial_{SS}\left[\sigma^2(S_t) f(S_t)\right] = \partial^2\sigma^2(S_t) f(S_t)/\partial^2 S_t$. We can get the stationary distribution by setting $\partial_t f(S_t) = 0$ as this tells us that the pdf at S_t does not change over time. Imposing this constraint, dropping the time subscript given that this is now a stationary distribution, and integrating once over S we get the current $J(S)$

$$J(S) = g(S) f(S) - \frac{1}{2}\partial_S\left[\sigma^2(S) f(S)\right].$$

As there is no baseline current due to deaths or births, just a reflecting lower bound, $J(S) = 0$. Using this, dividing by $\sigma^2(S_t) f(S_t)$, and rearranging, we have

$$\frac{\partial_S[\sigma^2(S) f(S)]}{\sigma^2(S) f(S)} = \partial_S \ln[\sigma^2(S) f(S)] = \frac{2g(S)}{\sigma^2(S)}.$$

Integrating this over S, we get an expression for the stationary distribution without frictions

$$f_0(S) = \frac{K}{\sigma^2(S)} \exp\left[2\int_S \frac{g(u)}{\sigma^2(u)} du\right],$$

where K is an integration constant ensuring that $\int_S f_0(u) du = 1$.

With a growth barrier When a growth barrier is in place, the forward Kolmogorov Equation (20) is the same as in the no friction case except in the slow growth interval $[\bar{S} - \delta, \bar{S}]$. Following a similar process as before for $f_0(S)$, but now separately for each interval $[S_{min}, \bar{S} - \delta[$, $[\bar{S} - \delta, \bar{S}]$, and $]\bar{S}, \infty]$, we have that[20].

$$f_1(S) = \begin{cases} K_1 f_0(S), & \text{if } S \in [S_{min}, \bar{S} - \delta[\\ K_2 \exp\left(-\frac{2\gamma}{\sigma^2}[S - (\bar{S} - \delta)]\right) f_0(S), & \text{if } S \in [\bar{S} - \delta, \bar{S}] \\ K_3 f_0(S), & \text{if } S \in]\bar{S}, \infty]. \end{cases}$$

Solving for $f_0(S)$ in each interval, we get

$$f_0(S) = \begin{cases} C_1 f_1(S), & \text{if } S \in [S_{min}, \bar{S} - \delta[\\ C_2 \exp\left(\frac{2\gamma}{\sigma^2}[S - (\bar{S} - \delta)]\right) f_1(S), & \text{if } S \in [\bar{S} - \delta, \bar{S}] \\ C_3 f_1(S), & \text{if } S \in]\bar{S}, \infty], \end{cases}$$

where $C_l = 1/K_l$. As $f_0(S)$ is continuous, we have that $C_1 = C_2$ and $C_3 = C_2 \exp(2\gamma\delta/\sigma^2)$. Given that $\int_S f_0(u)du = 1$, and δ is small, we have that

$$C_1(1 - F_1^+) + C_1 \exp(2\gamma\delta/\sigma^2) F_1^+ = 1.$$

Solving for C_1 and denoting $1 - p = \exp(-2\gamma\delta/\sigma^2)$, we have

$$C_1 = \frac{1 - p}{1 - p(1 - F_1^+)}$$

and hence

$$C_3 = \frac{1}{1 - p(1 - F_1^+)},$$

so that

$$f_0(S) = \begin{cases} \frac{1-p}{1-p(1-F_1^+)} f_1(S), & \text{if } S \in [S_{min}, \bar{S} - \delta[\\ \frac{1-p}{1-p(1-F_1^+)} \exp\left(\frac{2\gamma}{\sigma^2}[S - (\bar{S} - \delta)]\right) f_1(S), & \text{if } S \in [\bar{S} - \delta, \bar{S}] \\ \frac{1}{1-p(1-F_1^+)} f_1(S), & \text{if } S \in]\bar{S}, \infty] \end{cases}$$

[20] As δ is small, we can treat $\gamma(S)$ and $\sigma^2(S)$ as constant in the interval $[\bar{S} - \delta, \bar{S}]$.

Taking logs on both sides and approximating $\log(1-x) \approx -x$ for small x, we get

$$\log f_0(S) = \begin{cases} -pF_1^+ + \log f_1(S), & \text{if } S \in [S_{min}, \bar{S}-\delta[\\ -pF_1^+ + \frac{2\gamma}{\sigma^2}[S-(\bar{S}-\delta)] + \log f_1(S), & \text{if } S \in [\bar{S}-\delta, \bar{S}] \\ p(1-F_1^+) + \log f_1(S), & \text{if } S \in]\bar{S}, \infty], \end{cases}$$

which is our main result for the level shifts.

Why is p the probability that a firm that otherwise would have crossed the threshold does not do so? Let us concentrate on the interval $[\bar{S}-\delta, \bar{S}]$ and only those that enter this interval from below. Following a similar procedure as above, such firms have a pdf of $f_1(S)$ with the threshold, where

$$f_1(S) = K_2 \exp\left(-\frac{2\gamma}{\sigma^2}[S-(\bar{S}-\delta)]\right) f_0(S),$$

as above. However, as our counterfactual, we are now thinking of a situation where there is the same amount of probability mass coming in from below the threshold, as we are interested in what would happen to these same firms in different scenarios. Hence, we have that the incoming flow of firms from below is equal to D for both cases. Then, the counterfactual distribution has the form

$$\tilde{f}_0(S) = Lf_0(S).$$

Denoting by q the probability that a firm at the lower bound of the slow growth interval ($S = \bar{S}-\delta$) will exit the interval downward, we have a stationarity condition at the lower bound: $D = qf_1(\bar{S}-\delta) = L\tilde{f}_0(\bar{S}-\delta)$. Note that q is not affected by the slow growth region, so that we have $f_1(\bar{S}-\delta) = \tilde{f}_0(\bar{S}-\delta)$. This implies that $K = L$.

Now, denoting by l the probability that a firm at the threshold will exit the interval above, we have that the probability mass that has entered the slow growth interval from below and exits above is equal to $lh_0(\bar{S})$ without the threshold, and $lh_1(\bar{S})$ with the threshold, as l is not affected by the slow growth interval. This implies that the probability that a firm that enters the slow growth band from below exits above both with the threshold and without the threshold in place (i.e. $1-p$) is equal to

$$1-p = \frac{lf_1(\bar{S})}{l\tilde{f}_0(\bar{S})} = \frac{\exp\left(-\frac{2\gamma\delta}{\sigma^2}\right) f_0(\bar{S})}{f_0(\bar{S})} = \exp\left(\frac{-2\gamma\delta}{\sigma^2}\right).$$

Thus, p is the probability that a firm entering the slow growth band that would have exited it above without the threshold does not do so with the threshold in place.

C Appendix - Microfoundations

This appendix provides a more formal treatment of the microfoundations sketched in Section 2.3. We show that the drift reduction assumed in the mechanical framework of Section 2 arises from optimal firm behavior in a continuous-time model with a compliance cost above the policy threshold, and we characterize the quality of the approximation.

C.1 Setup

As in Section 2.1, firm size evolves according to the controlled Itô diffusion (see, e.g., Stokey (2009) for a textbook treatment)

$$dS_t = \big[g(S_t) + a_t\big]\, dt + \sigma(S_t)\, dW_t, \tag{21}$$

where $g(S)$ is the baseline drift, $\sigma(S)$ governs volatility, and $a_t \geq 0$ is the firm's growth effort (e.g., search for better business opportunities, investment, or innovation effort) that raises expected growth at a convex flow cost $c(a)$, with $c'(a) > 0$ for $a > 0$, $c'(0) = 0$, and $c'' > 0$.

Firms earn flow profits $\pi(S)$ and face a regulatory notch at $\bar{S}$ that imposes a fixed compliance cost $F > 0$ whenever the firm operates above the threshold:

$$\Pi(S) = \pi(S) - F \cdot \mathbf{1}\{S \geq \bar{S}\}. \tag{22}$$

Firms choose growth effort a_t to maximize the expected discounted value of future net payoffs with discount rate ρ:

$$V(S) = \max_{\{a_t \geq 0\}} \; \mathbb{E}_S\left[\int_0^\infty e^{-\rho t}\big[\Pi(S_t) - c(a_t)\big]\, dt\right]. \tag{23}$$

C.2 Value Functions

In the absence of the notch ($F = 0$), the value function $V_0(S)$ satisfies the Hamilton–Jacobi–Bellman equation:

$$\rho\, V_0(S) = \max_{a \geq 0} \left\{\pi(S) - c(a) + \big[g(S) + a\big]\, V_0'(S) + \tfrac{1}{2}\,\sigma^2(S)\, V_0''(S)\right\}. \tag{24}$$

With the notch in place, the value function $V_1(S)$ satisfies:

$$\rho\, V_1(S) = \max_{a \geq 0} \left\{\pi(S) - F \cdot \mathbf{1}\{S \geq \bar{S}\} - c(a) + \big[g(S) + a\big]\, V_1'(S) + \tfrac{1}{2}\,\sigma^2(S)\, V_1''(S)\right\}, \tag{25}$$

with the continuity conditions $V_1(\bar{S}^-) = V_1(\bar{S}^+)$ and $V_1'(\bar{S}^-) = V_1'(\bar{S}^+)$. Because F enters as a flow cost (rather than a one-off lump-sum payment upon crossing), there is no jump in the value function at $\bar{S}$. Smooth pasting of V_1' also holds under standard regularity conditions (see Dixit (1993)), while V_1'' may have a kink at $\bar{S}$ to accommodate the discontinuity in $F \cdot \mathbf{1}\{S \geq \bar{S}\}$.

In both regimes, the first-order condition for effort is

$$c'(a^*) = V'(S), \tag{26}$$

with $V' \in \{V_0', V_1'\}$. Strict convexity of c implies that optimal effort $a^*(S)$ is strictly increasing in $V'(S)$.

C.3 Drift Reduction Below the Threshold

The compliance cost F above $\bar{S}$ reduces the value function: $V_1(S) < V_0(S)$ for all S. The effect on the slope differs by region.

Far below the threshold. For $S \ll \bar{S}$, the probability of reaching $\bar{S}$ in any finite horizon is negligible, so $V_1(S) \approx V_0(S)$ and $V_1'(S) \approx V_0'(S)$. This implies that effort and hence drift are essentially unaffected.

Near the threshold from below. For some interval $S \in [\bar{S} - \delta, \bar{S}]$, the firm faces a non-trivial probability of crossing $\bar{S}$ and incurring the compliance cost F. This reduces the marginal value of growing toward the threshold:

$$V_1'(S) < V_0'(S) \quad \text{for } S \in [\bar{S} - \delta, \bar{S}], \tag{27}$$

where $\delta > 0$ depends on σ, ρ, and F. By (26), this implies

$$a_1^*(S) < a_0^*(S) \quad \text{for } S \in [\bar{S} - \delta, \bar{S}], \tag{28}$$

i.e., optimal growth effort and hence the drift of the size process are lower in this region when the notch is in place.

C.4 Behavior Above the Threshold

The compliance cost also affects the value function above $\bar{S}$, so $V_1(S) < V_0(S)$ for $S > \bar{S}$ as well. Whether this translates into a drift distortion depends on how the gap $\Delta V(S) \equiv$

$V_0(S) - V_1(S)$ varies with S.

Far above the threshold. For $S \gg \bar{S}$, the firm is almost always above $\bar{S}$, and F acts as a simple reduction in flow profits. The gap satisfies

$$\Delta V(S) \approx \frac{F}{\rho} \quad \text{(constant)}, \tag{29}$$

so that $\Delta V'(S) \approx 0$ and hence

$$V_1'(S) \approx V_0'(S) \quad \text{for } S \gg \bar{S}. \tag{30}$$

Effort and drift are approximately unchanged.

To see (29) more formally, subtract (25) from (24). At the respective optima a_0^* and a_1^*, and for $S > \bar{S}$, the gap $\Delta V = V_0 - V_1$ satisfies

$$\rho\,\Delta V(S) = F + \big[c(a_1^*) - c(a_0^*)\big] + g(S)\,\Delta V'(S) + a_0^*\,V_0'(S) - a_1^*\,V_1'(S) + \tfrac{1}{2}\,\sigma^2(S)\,\Delta V''(S). \tag{31}$$

When $\Delta V' \approx 0$ and $\Delta V'' \approx 0$ (as is the case far from $\bar{S}$), and $a_0^* \approx a_1^*$ (which follows from $V_0' \approx V_1'$), the equation reduces to $\rho\,\Delta V \approx F$, confirming (29).

Near the threshold from above. At $S = \bar{S}^+$, the gap ΔV is smaller than F/ρ: a firm just above $\bar{S}$ may diffuse back below and temporarily avoid the compliance cost, reducing the effective present value of F. Consequently, $\Delta V'(\bar{S}^+) > 0$ and $V_1'(\bar{S}^+) < V_0'(\bar{S}^+)$. There is a small, localized drift reduction just above the threshold. This effect decays as S moves away from $\bar{S}$.

C.5 Mapping to the Mechanical Framework

Define the drift reduction as in Section 2.3:

$$\gamma(S) \equiv a_0^*(S) - a_1^*(S) \geq 0. \tag{32}$$

The endogenous drift with the notch in place is then

$$g_1(S) \equiv g(S) + a_1^*(S) = \underbrace{\big[g(S) + a_0^*(S)\big]}_{g_0(S)} - \gamma(S), \tag{33}$$

where $g_0(S)$ and $g_1(S)$ denote the endogenous drifts without and with the notch, respectively. The results above establish that:

- $\gamma(S) \approx 0$ for $S \ll \bar{S}$ (threshold too distant to affect incentives).
- $\gamma(S) > 0$ for $S \in [\bar{S} - \delta, \bar{S}]$ (anticipated compliance cost reduces growth effort).
- $\gamma(S) \approx 0$ for $S \gg \bar{S}$ (cost is position-independent, shifting the level of V_1 but not its slope).
- $\gamma(S)$ is small but positive in a neighborhood above $\bar{S}$ (option value of diffusing back below the threshold).

The mechanical framework in Equation (5) approximates this by setting $\gamma > 0$ only in $[\bar{S} - \delta, \bar{S}]$ and $\gamma = 0$ everywhere else:

$$g_1(S) \approx \begin{cases} g_0(S), & S < \bar{S} - \delta, \\ g_0(S) - \gamma, & S \in [\bar{S} - \delta, \bar{S}], \\ g_0(S), & S > \bar{S}, \end{cases}$$

which corresponds directly to Equation (8). This approximation is accurate when F is small relative to $\pi(\bar{S})$, so that the above-threshold drift distortion is second-order, and when δ is small relative to the scale of the distribution, so that the step-function approximation to the smooth $\gamma(S)$ is adequate.

Direction of the approximation error. The small drift reduction just above $\bar{S}$ that the mechanical model ignores works in the direction of reinforcing the growth barrier. Firms just above the threshold grow slightly more slowly than the mechanical model assumes, increasing the probability of falling back below $\bar{S}$. If anything, this would amplify the distributional distortions predicted by the mechanical model: more mass would accumulate below the threshold and less above it than the model predicts. The counterfactual distribution constructed from the mechanical framework is therefore conservative.

D Appendix – Switcher DiD Identification

This appendix shows that a standard ATET is point-identified under a set of assumptions when the notch causes some units to constrain their size below the threshold. We derive a closed-form expression that decomposes the ATET into a standard difference-in-differences estimator plus a correction term that captures the bias arising from switchers.

Setup

We consider a setting with two periods: a pre-reform period $t = 0$ in which a policy notch is in place, and a post-reform period $t = 1$ in which the notch has been repealed. The realized outcome of unit f in period t is $y_{f,t}$. In each period, units belong to one of three latent groups:

1. *Below-stayers* (bs): units that would locate below the notch regardless of whether the notch is in place.

2. *Above-stayers* (as): units that would locate above the notch regardless of whether the notch is in place.

3. *Switchers* (sw): units that locate below the notch when it is in place, but would locate above the notch in the absence of the notch.

Group identifiers for these different types of units are $g \in \{bs, as, sw\}$ respectively. The number of units in group g in period t is n_t^g, and the average outcome of the group is $y_t^g \equiv E(y_{f,t} \mid f \in g \text{ during } t)$. Switchers and above-stayers are defined as treated by the reform, as they would have located above the notch in the absence of the notch. Below-stayers are non-treated.

We use potential outcome notation, so that $y_t^g(R) = E(y_f(R) \mid f \in g \text{ during } t)$ is the potential expected outcome of group g in period t when the notch is either still in place ($R = 0$) or has been repealed ($R = 1$). We observe the number of units above ($a = 1$) and below ($a = 0$) the threshold, $n_{a,t}$, and the corresponding mean outcomes $y_{a,t}$. The definitions of the groups imply:

$$\begin{cases} n_{0,0} = n_0^{bs} + n_0^{sw}, \\ n_{1,0} = n_0^{as}, \\ n_{0,1} = n_1^{bs}, \\ n_{1,1} = n_1^{as} + n_1^{sw}, \end{cases} \tag{34}$$

and the observed mean outcomes satisfy:

$$\begin{cases} y_{0,0} = y_0^{bs} \dfrac{n_0^{bs}}{n_0^{bs} + n_0^{sw}} + y_0^{sw} \dfrac{n_0^{sw}}{n_0^{bs} + n_0^{sw}}, \\ y_{1,0} = y_0^{as}, \\ y_{0,1} = y_1^{bs}, \\ y_{1,1} = y_1^{as} \dfrac{n_1^{as}}{n_1^{as} + n_1^{sw}} + y_1^{sw} \dfrac{n_1^{sw}}{n_1^{as} + n_1^{sw}}. \end{cases} \tag{35}$$

Target parameter

Our target parameter is the ATET in the post-reform period $t = 1$, which can be expressed as

$$\begin{aligned} \text{ATET} &= E\big(y_{f,1}(1) \mid f \text{ is treated}\big) - E\big(y_{f,1}(0) \mid f \text{ is treated}\big) \\ &= E\big(y_1^g(1) \mid g \in \{as, sw\}\big) - E\big(y_1^g(0) \mid g \in \{as, sw\}\big). \end{aligned}$$

Assumptions

To identify ATET, we make the following assumptions:

Assumption 1 (Parallel trends in log numbers of units)**.**

$$\log n_1^{bs} - \log n_0^{bs} = \log n_1^{as} - \log n_0^{as} = \log n_1^{sw} - \log n_0^{sw}.$$

Assumption 2 (Parallel trends in level mean outcomes)**.**

$$y_1^{bs}(0) - y_0^{bs}(0) = y_1^{as}(0) - y_0^{as}(0) = y_1^{sw}(0) - y_0^{sw}(0).$$

Assumption 3 (Potential outcomes of switchers without the reform)**.**

$$y_t^{sw}(0) = y_t^{bs}.$$

Assumption 4 (Below-stayers not affected by the reform)**.**

$$y_t^{bs}(0) = y_t^{bs}(1).$$

Identification of ATET

We show that under Assumptions 1–4, ATET can be expressed as

$$\text{ATET} = \underbrace{\tau^{DID}}_{\text{Diff-in-diff estimate}} + \underbrace{w(y_{1,0} - y_{0,0})}_{\text{Switcher term}},$$

where $\tau^{DID} = (y_{1,1} - y_{0,1}) - (y_{1,0} - y_{0,0})$ is the standard difference-in-differences estimate comparing means post- and pre-reform below and above the notch point, and w is the share of switchers among treated units in the post-reform period.

There are two treated groups: the above-stayers and the switchers. The ATET is then given by

$$\text{ATET} = \left(y_1^{as}(1) \tfrac{n_1^{as}}{n_1^{as}+n_1^{sw}} + y_1^{sw}(1) \tfrac{n_1^{sw}}{n_1^{as}+n_1^{sw}} \right) - \left(y_1^{as}(0) \tfrac{n_1^{as}}{n_1^{as}+n_1^{sw}} + y_1^{sw}(0) \tfrac{n_1^{sw}}{n_1^{as}+n_1^{sw}} \right). \tag{36}$$

Note the expression for $y_{1,1}$ from (35) and that $y_1^{sw}(1) = y_1^{sw}$ and $y_1^{as}(1) = y_1^{as}$, so that the expression for ATET in (36) can be written as

$$\text{ATET} = y_{1,1} - y_1^{as}(0) \frac{n_1^{as}}{n_{1,1}} - y_1^{sw}(0) \frac{n_1^{sw}}{n_{1,1}}. \tag{37}$$

Now, the parallel trends assumption (Assumption 2) implies that $y_1^{as}(0) - y_0^{as}(0) = y_1^{sw}(0) - y_0^{sw}(0) = y_1^{bs}(0) - y_0^{bs}(0)$. Combining this with (35) and Assumptions 3 and 4 we have that $y_1^{bs}(0) - y_0^{bs}(0) = y_{0,1} - y_{0,0}$, as in the pre-period, the mean outcome below the threshold is equal to the mean outcome of below-stayers (as switchers have the same mean outcome by Assumption 3), and in the post-period, the mean outcome below the threshold is equal to the mean outcome of below-stayers simply because they are the only ones there, and as by Assumption 4, the below-stayers are not affected by the reform.

Using (35) we have $y_0^{as}(0) = y_{1,0}$ and by Assumption 3 $y_0^{sw}(0) = y_0^{bs} = y_{0,0}$. Therefore,

$$y_1^{as}(0) = y_{1,0} + y_{0,1} - y_{0,0}, \tag{38}$$

$$y_1^{sw}(0) = y_{0,0} + y_{0,1} - y_{0,0} = y_{0,1}. \tag{39}$$

Substituting (38) and (39) into (37):

$$
\begin{aligned}
\text{ATET} &= y_{1,1} - (y_{1,0} + y_{0,1} - y_{0,0})\frac{n_1^{as}}{n_{1,1}} - y_{0,1}\frac{n_1^{sw}}{n_{1,1}} \\
&= y_{1,1} - y_{1,0}\frac{n_1^{as}}{n_{1,1}} - y_{0,1} + y_{0,0}\frac{n_1^{as}}{n_{1,1}} \\
&= \underbrace{(y_{1,1} - y_{0,1}) - (y_{1,0} - y_{0,0})}_{\tau^{DID}} + \frac{n_1^{sw}}{n_{1,1}}(y_{1,0} - y_{0,0}),
\end{aligned}
$$

where the last line follows from $n_1^{sw}/n_{1,1} = 1 - n_1^{as}/n_{1,1}$. Hence, we have that

$$
\text{ATET} = \underbrace{\tau^{DID}}_{\text{Diff-in-diff estimate}} + \underbrace{\frac{n_1^{sw}}{n_{1,1}}(y_{1,0} - y_{0,0})}_{\text{Switcher term}}.
$$

It remains to show that the switcher share $w \equiv n_1^{sw}/n_{1,1}$ is identified from observables. From (34), $n_1^{sw} = n_{1,1} - n_1^{as}$. Moreover, the assumption on parallel log trends in the number of firms of each group (Assumption 1) implies that the total — and observable — number of firms in the economy $n_t = n_t^{as} + n_t^{bs} + n_t^{sw}$ has the same trend as all sub-groups $g \in \{bs, as, sw\}$. Hence, we have that $n_1^{as}/n_0^{as} = n_1/n_0$. Also, note from (34) that $n_0^{as} = n_{1,0}$. Defining F_{1-R}^{+} as the share of firms above the threshold when the notch is either still in place ($R = 0$) or has been repealed ($R = 1$)[21], w can be expressed as

$$
w = 1 - \frac{F_1^+}{F_0^+},
$$

and all quantities on the right-hand side are observed. Hence, the ATET becomes

$$
\text{ATET} = \tau^{DID} + w(y_{1,0} - y_{0,0}).
$$

[21] This is the same object as in our diffusion theory in Section 2.

E Appendix – Are Firm Distributional Responses in Section 4.3 Real?

A potential explanation for firm distributional responses in Section 4.3 is that firms may engage in tax avoidance or evasion strategies to remain below the threshold. However, we provide several pieces of evidence indicating that this is not the case. First, Panel (a) of Figure A16 shows a very similar firm distribution when using depreciation reported in accounting data, as in Figure 2b. This accounting-based measure is less prone to manipulation than the corresponding tax-based measure, as approximately 90% of firms in our sample are required to have their accounts audited by a third party. Second, firm owners and entrepreneurs might attempt to avoid the threshold by splitting their firms. Yet, as shown in Panel (b) of Figure A16, we find no evidence of such behavior – the number of firms per owner remains smooth below and above the threshold. Although there appears to be a slight difference at higher levels of depreciation, this difference is not statistically significant for most bins (as illustrated in Panel (d) of Figure A16). Finally, we conduct an additional test for potential firm-splitting by allocating depreciation to owners exploiting our matched owner-firm data. If avoidance were present, we would observe a discontinuity in this distribution at the threshold. As shown in Panel (c) of Figure A16, no such discontinuity is evident.

Another possible form of firm-level avoidance behavior would be to substitute rental capital for owned capital in production, thereby reducing the level of reported depreciation. While we do not have a precise measure for rents specifically to capital usage, we do observe aggregate rental costs at the firm level. Appendix A Panel (a) of Figure A17 shows the evolution of firm-level rental payments, which appear to develop smoothly below the threshold during the period when the tax notch was in place. However, the slopes above the threshold differ between the pre- and post-2010 regimes, suggesting that rental payments declined after the repeal of the tax notch relative to the pre-repeal period among firms above the threshold. This pattern is particularly evident in Panel (b) of Figure A17, which plots the difference in rental payments between the pre- and post-2010 periods across depreciation bins. It is worth noting, however, that rental payments represent only a small fraction of total capital depreciation on average, indicating that this behavioral response is unlikely to be driving the overall effects observed around the threshold.

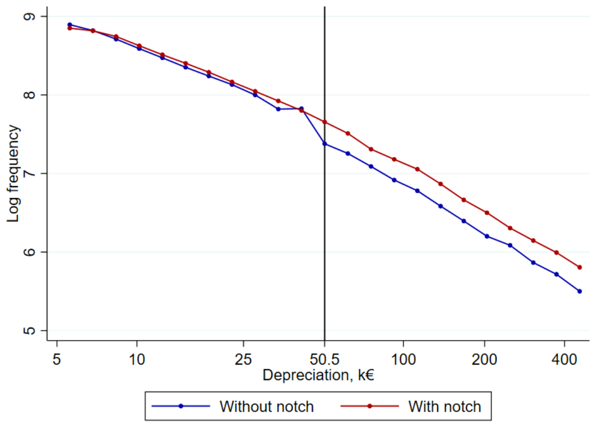
Firms above the threshold might also attempt to avoid payroll taxes by outsourcing part of their workforce. Although we do not have a precise measure of the share of outsourced workers, we observe firm-level spending on services purchased from outside providers, which captures outsourcing costs among other related expenses. Panel (c) of Appendix A, Figure

A17, shows no discontinuity at the threshold in services purchased externally during the period when the threshold induced a payroll tax rate jump, nor any change at the threshold after its repeal. However, among firms with approximately €150,000 in capital depreciation, we observe a decline in the difference measure, suggesting that some firms further away from the threshold may have responded to the payroll tax increase by increasing their reliance on outsourced labor.

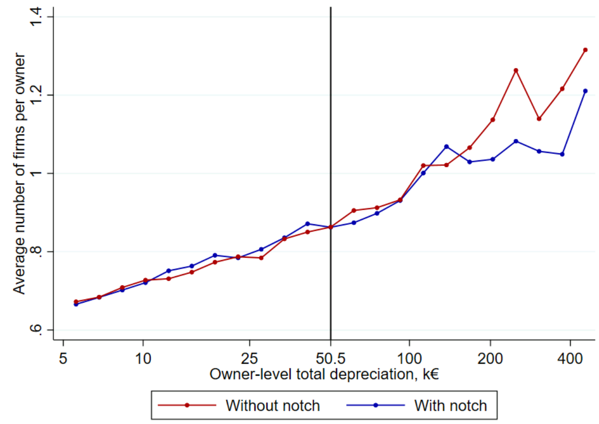
Overall, our results suggest that a firm-level payroll tax notch linked to annual capital depreciation generates substantial changes in the firm-size distribution. The empirical evidence is inconsistent with explanations based on tax avoidance or evasion.These findings validate the theoretical framework from Section 2 and motivate the methodological approach in Section 3: the threshold creates a growth barrier that induces firm movement across the distribution and affects exit decisions.

Two potential mechanisms drive these shifts in the firm distribution. First, higher payroll taxes may reduce the resilience of firms above the threshold to adverse shocks, leading to higher exit rates among larger firms (exit-switchers). Second, the notch may discourage firms from growing beyond the threshold, since even a small increase in capital depreciation would trigger a substantial rise in payroll tax liability (notch-switchers). In a standard bunching framework (Kleven, 2016), such behavior would not by itself produce a large decline in the number of firms far above the threshold. However, if there are for example learning-by-doing dynamics, where firms must first learn to operate at a slightly lower scale before expanding further, a notch could also lead to fewer firms above the threshold even in the absence of extensive-margin effects.

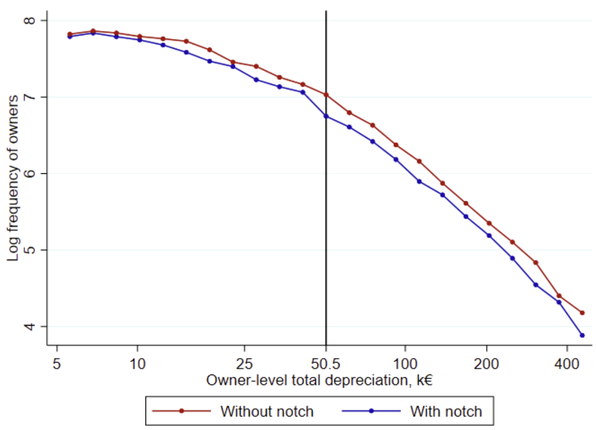
(a) Tax Depreciation in Accounting



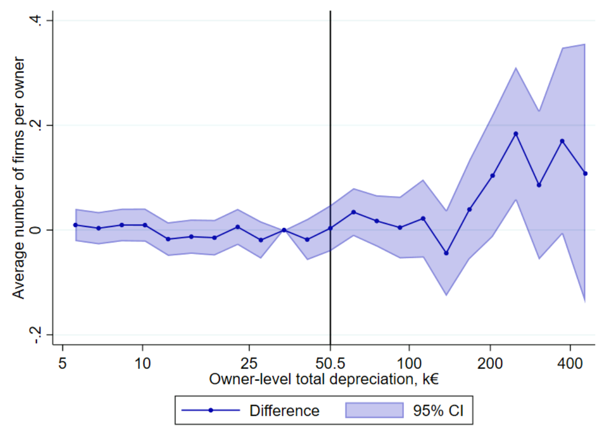
(b) Number of firms per owner



(c) Log distribution of owners

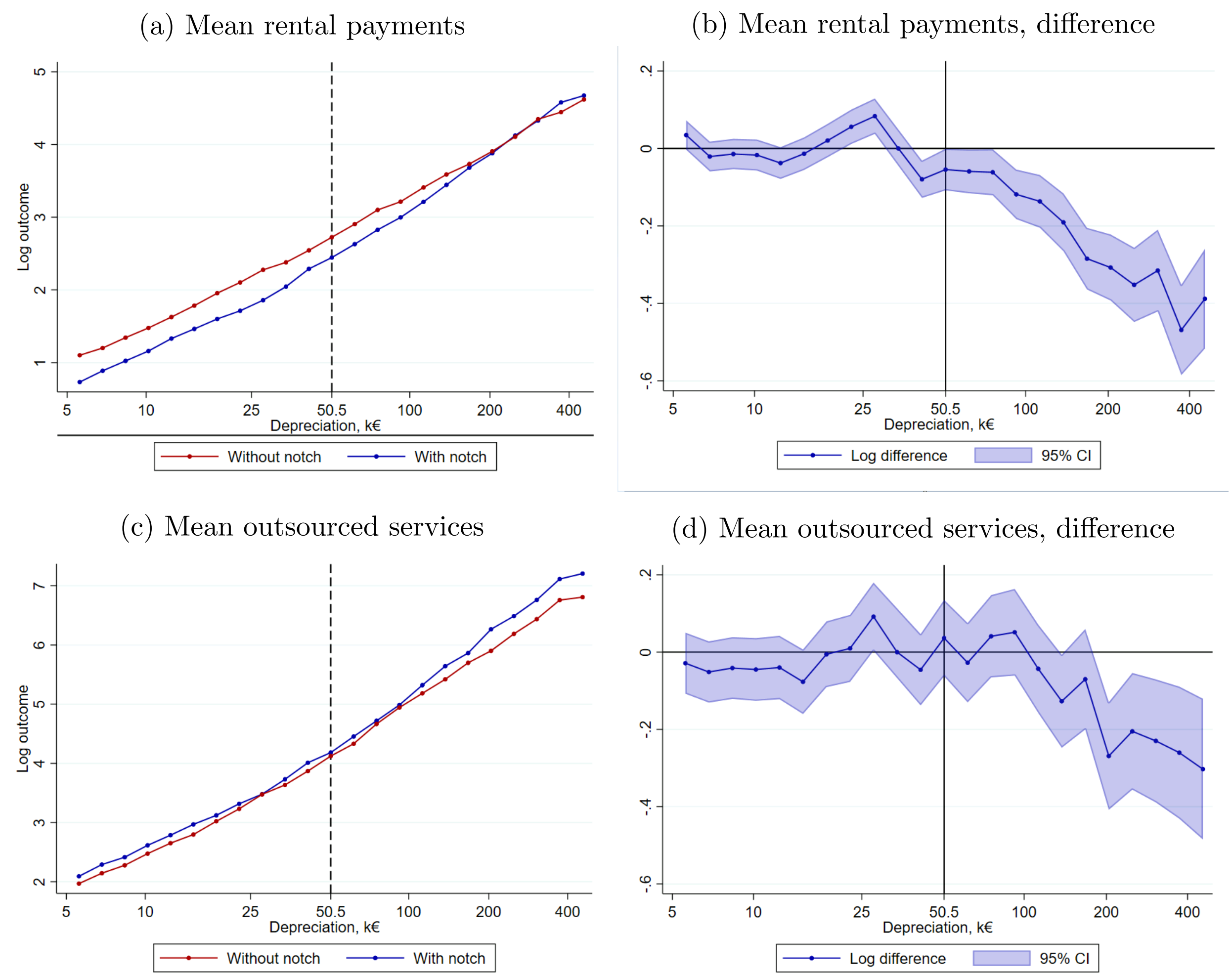


(d) Number of firms per owner, difference



Notes: Figure A16a plots the log frequency of firms using the tax measure of capital depreciation as a running variable with (2005–2007) and without notch (2011–2013). The panel A16b shows the mean number of firms per owner within each bin with (2003–2007) and without notch (2011–2015), and the panel A16d depicts the difference of these with 95% confidence intervals. Panel A16c shows the log number of owners across the owner-level depreciation distribution. The width of the bins is calculated such that each bin captures 20 log points of capital depreciation up and down starting from the threshold. Only firms/owners that had capital depreciation more than 10% of their labor costs are included as the discontinuity only applied to such firms (see more details on the institutional setup in Section 4.1)

Figure A16: Evidence of Avoidance and Evasion Responses



Notes: The panel A17a shows the mean of rental payments within each bin with (2003–2007) and without notch (2011–2015), and the panel A17b depicts the difference of these two regimes with 95% confidence intervals. The panel A17c shows the mean of outsourcing within each bin with (2003–2007) and without notch (2011–2015), and the panel A17d depicts the difference of these two regimes with 95% confidence intervals. The width of the bins is calculated such that each bin captures 20 log points of capital depreciation up and down starting from the threshold.

Figure A17: Firms using outsourcing or rental capital to avoid exceeding the threshold

F Methodology for Aggregate Effects and the Fiscal Externality

The firm-level treatment effects translate into aggregate quantity changes and a fiscal externality. Because the parallel-trends assumptions hold in logs, we estimate effects in logs and aggregate in levels by applying each subgroup's log effect to its own base.

The overall log ATET in Equation (10) is a count-weighted average of the effects on above-stayers (as) and switchers (sw),

$$\mathrm{ATET} = (1-\hat{w})\,\mathrm{ATET}^{as} + \hat{w}\,\mathrm{ATET}^{sw}, \qquad \hat{w} = 1 - \frac{F_1^+}{F_0^+}. \tag{40}$$

This identity does not separate the two effects. We pin the split by assuming that switchers and above-stayers have equal mean log outcomes once both are above the threshold post-reform, $y_1^{sw}(1) = y_1^{as}(1)$. The above-stayer effect is then the conventional difference-in-differences estimate and the switcher effect is that plus the pre-reform cross-threshold gap,

$$\mathrm{ATET}^{as} = \hat{\delta}, \qquad \mathrm{ATET}^{sw} = \hat{\delta} + \hat{\beta}, \tag{41}$$

with $\hat{\delta} = \tau^{DID}$ and $\hat{\beta} = y_{1,0} - y_{0,0}$ from Equation (11).

For outcome k, the aggregate level change applies each subgroup's effect to its pre-reform base,

$$\Delta X_k = X_{k,0}^{sw}\big(e^{\,\mathrm{ATET}_k^{sw}} - 1\big) + X_{k,0}^{as}\big(e^{\,\mathrm{ATET}_k^{as}} - 1\big), \tag{42}$$

where $X_{k,0}^{as}$ is the pre-reform total among above-threshold firms and $X_{k,0}^{sw} = \phi \sum_{\text{below},0} X_k$ is the switcher base, with ϕ the switcher share among below-threshold firms in the pre-period. Assumption 3 places switchers at the below-stayer mean, so scaling the below-threshold total by ϕ recovers their base. Exponentiation matters for the switcher term, whose effect is large.

The fiscal externality compares the behavioral revenue change to the mechanical cost of the rate cut. Only the employer payroll rate changed, and before the reform only above-stayers paid the surtax, so the mechanical cost is $M = \Delta\tau_p\, W_0^{as}$, with $\Delta\tau_p \approx 0.028$ and W_0^{as} the pre-reform payroll base of above-threshold firms. The behavioral change sums statutory rates times the aggregate base changes from (42),

$$B = \tau_L\,\Delta(\text{Labor costs}) + \tau_v\,\Delta(\text{Value added}) + \tau_c\,\Delta(\text{Profits}), \tag{43}$$

with $\tau_L = 0.423$ the OECD labor tax wedge for an average single worker in Finland in 2010, $\tau_v = 0.22$ the value-added tax rate, and $\tau_c = 0.26$ the corporate rate. The full wedge absorbs employer payroll, employee social contributions, and income tax in one term. The fiscal externality in terms of the mechanical effect is $\mathrm{FE} = B/M$, and the naive counterpart replaces each subgroup effect with the uncorrected $\hat{\delta}_k$.