Estimating the Effects of Size-Based Regulation: A New Approach and Evidence from a Firm Payroll Tax Notch*

Youssef Benzarti† Jarkko Harju‡ Sami Jysmä§

February 9, 2026
Please do not distribute.

Abstract

Discrete policy thresholds are widespread in tax and regulatory systems and can substantially affect firm and individual behavior. This paper first develops a theoretical framework showing that when policy notches act as barriers to mobility within a distribution, such as "growth barriers" for firms or individual "welfare cliffs", they generate distortions that extend far beyond the threshold itself. We then demonstrate that conventional difference-in-differences estimators yield biased estimates in such cases and propose a new methodology that offers bounds on causal effects. Finally, we apply the framework to study the abolition of a payroll tax notch in Finland, under which firms faced a 2.5 percentage point increase in the tax rate crossing a capital depreciation threshold. We find that the notch acts as a significant growth barrier, reducing the number of firms above the threshold by approximately 18 percent, with effects extending far above the cutoff. The notch also systematically reduced firm scale, including employment, capital stock and value added, by approximately 10 percent.

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

*We are grateful to the Academy of Finland (grant no. 346252) for funding for this project.
†University of Santa Barbara & NBER.
‡Finnish Centre of Excellence in Tax Systems Research (FIT) & Tampere University & CESifo.
§Labore & Tampere University & FIT.

1 Introduction

Discrete policy thresholds are pervasive in tax and regulatory systems. Tax notches impose discontinuous jumps in liability when income or size exceeds a cutoff. Welfare cliffs terminate benefits at eligibility limits for individuals. Employment, tax, environmental and social regulations exempt firms below specified thresholds, imposing discrete compliance costs on those above. In each case, crossing a threshold triggers a discontinuous change in incentives, a policy notch.

Understanding how such discrete thresholds shape behavior is central to policy design. A large literature has documented bunching responses (see Kleven (2016)), where individuals or firms concentrate just below thresholds to avoid triggering discontinuous costs. Yet these settings tend to consider static within year behavior and thus capture only local responses around the threshold. In dynamic environments by contrast discrete thresholds can affect whether individuals or firms cross the threshold at all and can reshape the entire distribution of outcomes, with aggregate effects that far exceed those implied by local bunching estimates. Large distributional responses of this kind have been documented, for example, in the context of firm regulation (e.g. Garicano et al. 2016; Gourio and Roys 2014; Aghion et al. 2023b) and individual tax reporting requirements ((Garbinti et al., 2023)), but we lack a general and portable framework to study these impacts in different settings.

This paper studies the distortions generated by discrete policy thresholds and develops new theoretical and empirical tools to quantify their effects. We show in a theoretical framework that when policy notches act as barriers to mobility within a distribution, they can generate distortions that extend far beyond the threshold itself. We further demonstrate that conventional difference-in-differences estimators are biased in such settings and propose an alternative approach that can identify causal effects. We apply these tools to a firm-level payroll tax notch in Finland, exploiting the abolition of a size-based threshold as a natural experiment. The empirical results indicate that the notch constituted a significant growth barrier, with effects extending well above the cutoff, and that it reduced firm scale by nearly 10 percent.

Our analysis proceeds in three parts, each offering a distinct contribution.

A theoretical framework for notch-induced distributional distortions. We develop a mechanistic but portable model in which a discrete policy threshold acts as a barrier to mobility within a distribution. When crossing a threshold triggers a discontinuous increase in costs or a loss of benefits, firms or individuals have strong incentives to remain below it. We show that for a broad class of stochastic processes governing mobility, including the proportional random growth processes widely used to model firm dynamics, earnings

evolution, and wealth accumulation (see e.g. Gabaix (2009); Benhabib and Bisin (2018)), such barriers generate systematic and predictable distortions in the stationary distribution. Crucially, these distortions are not confined to a narrow region around the threshold. Instead, by discouraging crossing, the barrier reduces mass throughout the region above the threshold while increasing mass below it.

Similar mechanisms have previously been included in more structural models of firm dynamics (e.g. Gourio and Roys (2014); Aghion et al. (2023b)). Our contribution is to show that the effects of threshold-induced distortions can be summarized by relatively simple sufficient statistics that link the observed distorted distributions to the counterfactual distribution that would prevail in the absence of the threshold. The crucial sufficient statistic is the probability that an individual or a firm refrains from crossing the threshold when the notch is in place. We show that this probability can be quasi-experimentally identified using data from periods in which the notch is active, by adapting the dynamic bunching approach of Garbinti et al. (2023).

This framework yields a sharp empirical prediction: removing the threshold should lead to an increase in the mass above the former cutoff, with a magnitude determined by the identifiable probability of refraining from crossing. Evidence from our empirical setting strongly supports this prediction, as the estimated counterfactual firm distribution matches the observed undistorted distribution after the notch was repealed.

While our application is for firms, the theoretical results apply broadly to settings in which individuals or firms evolve according to diffusion processes and face discrete threshold-based incentives, including, for example, notched tax schedules, welfare cliffs in transfer programs, and regulatory size thresholds.

A new methodology for estimating treatment effects with notch-induced distributional changes. Standard difference-in-differences methods assume that treatment status is observed and fixed. This assumptions fails when discrete policy thresholds induce units to relocate across cutoffs or to exit the distribution entirely, as in our theoretical framework. In such settings, firms or individuals that refrain from crossing the threshold are treated by the policy, but a conventional difference-in-differences strategy comparing units below and above the threshold would classify them as untreated. This misclassifications contaminates the control group and leads to biased estimates.

We address this identification challenge by developing estimators that bound an informative log average treatment effect on the treated under alternative assumptions about the nature of movement in the distribution. This methodological development contributes to the rapidly expanding literature on difference-in-differences (see e.g. Abadie et al. (2025); Roth et al. (2023); de Chaisemartin and D'Haultfœuille (2023) for recent surveys of this

literature).

This proposed methodology has many applications. Any discrete threshold-based policy that generates distributional shifts raises similar identification challenges. Examples include welfare cliffs affecting labor supply and savings decisions, asset tests affecting wealth accumulation, and size-based regulations shaping firm growth. Our bounding approach provides a tractable solution in settings where standard methods fail due to policy-induced movement across thresholds.

Empirical evidence on firm responses and incidence. Applying our methodology to rich administrative data linking Finnish employers to their employees and owners. Studying the abolition of a firm-level payroll tax notch we document four sets of findings.

We first document large impacts on the firm distribution and dynamics that extend well beyond the threshold. When the payroll tax notch was in place, we observe pronounced bunching just below it and, more strikingly, a persistent 20 percent reduction in mass above it, extending to firms with depreciation eight times the threshold level. We provide evidence that these responses reflect genuine changes in firm scale rather than avoidance through reporting, firm splitting, outsourcing, or renting capital goods. Following the repeal of the threshold, the firm distribution by capital depreciation is smooth and closely aligns with the "power law" regularity observed in many countries, including the United States ((Axtell, 2001; Gabaix, 2016)).

These results extend existing research on the effects of size-based thresholds in taxation and regulation. Earlier work has documented sharp behavioral responses around thresholds, including bunching below the turnover and profit tax cutoffs in Pakistan ((Best et al., 2015)), responses to size-contingent labor regulations in France ((Gourio and Roys, 2014; Aghion et al., 2023b; Garicano et al., 2016)), tax enforcement threshold in Spain ((Almunia and Lopez-Rodriguez, 2018)), and behavioral responses to VAT thresholds in Japan and Finland ((Onji, 2009; Harju et al., 2019)). By studying the abolition of a payroll tax threshold in Finland, we provide new evidence on how size-dependent tax policies can generate wide-ranging distortions in firm dynamics that extend far beyond local responses at the cutoff.

Second, the repeal of the notch led to substantial increases in firm scale. Labor costs rose by 8.5–13.3 log points, employment by 4.8–10.3 log points, and capital stock by 9.8–15.1 log points, with comparable increases in revenue and value added. These estimates are robust to using unaffected employers (those below the capital intensity threshold that determined eligibility) as an additional control group in a triple-difference specification. Overall, these results are consistent with empirical evidence on payroll taxes and firm scale ((Saez et al., 2019; Cottet, 2023)) and extend earlier work by Benzarti and Harju (2021b) using the same reform by moving beyond local responses at the threshold to a broader set of firms using our

new approach.

Third, we examine how payroll taxes affect input composition. Following the tax reduction, affected employers increased their use of non-routine and highly educated workers, suggesting that payroll taxes may be particularly distortionary for high-skilled labor. On the capital side, investments shift away from buildings toward research and development. These findings contribute to a growing literature studying how policy-induced changes in factor costs and incentives affect firm-level production technologies, worker composition and investment decisions ((Criscuolo et al., 2019; Aghion et al., 2023a; Hirvonen et al., 2025; Curtis et al., 2025; Lichter et al., 2025)).

Finally, we estimate payroll tax incidence. Decomposing changes in value added into components accruing to labor versus capital, we find that workers bear approximately 63 percent of the tax burden and owners bear 37 percent. Importantly, the worker-side incidence operates almost entirely through the employment margin: average employee-level wages respond little, whereas employment adjusts substantially. This contrasts with traditional approaches emphasizing wage pass-through as the primary channel of labor tax incidence. Our results contribute to a growing body of work on the incidence of payroll taxes (e.g., Gruber 1997; Korkeamäki and Uusitalo 2009; Saez et al. 2012, 2019; Bozio et al. 2023; Cottet 2023; Benzarti and Harju 2021a; Benzarti 2025).

The remainder of the paper proceeds as follows. Section 2 develops the theoretical framework showing how policy thresholds can act as growth barriers. Section 3 presents our estimation methodology for quantifying treatment effects when thresholds induce distributional shifts. Section 4 describes the Finnish institutional setting and data. Section 5 tests the theoretical predictions and validates our methodological assumptions using the Finnish payroll tax reform. Section 6 presents the main empirical results on firm scale, input demand, and tax incidence. Section 7 concludes.

2 Theoretical Framework

In this section, we show that size-based policy thresholds, above which regulations, financial or tax reporting requirements, or fixed costs such as taxes increase sharply, can generate substantial distortions in the distributions of affected units, 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 growth rates of firms or the growth rates of incomes or wealth of individuals just below it. The key implication of this theoretical result is that

when a policy threshold acts as a growth barrier, it can significantly reshape the overall distribution. This, in turn, may have important welfare consequences: in the case of firms, aggregate value added and total employment may be considerably lower in the presence of the threshold than they otherwise would be. Or, in the case of individuals, aggregate earnings or wealth may be substantially lower.

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 position of the process. In our setting, this would mean that growth rates and growth rate volatility could depend on unit size:

$$dS_t = g(S_t)dt + \sigma(S_t)dW_t, \tag{1}$$

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 that there is a minimum size for firms (modeled as a reflecting wall, see Gabaix 2009) at $S_{\min}$. Another alternative would be, for example, to assume a suitable Poisson exit rate for units. Note that a random proportional growth process (or geometric Brownian motion) is a special case of the process in Equation (1) such that $g(S_t) = gS_t$ and $\sigma(S_t) = \sigma S_t$. Proportional random growth processes have been widely used in the firm distribution literature as they produce a power law distribution under small frictions that fits many observed distributions well (see e.g. Gabaix 2009 for a review). However, Equation (1) is a more general diffusion process than most of the processes used in this literature as it allows for state dependent relative drift and variance.

We study a case where the government introduces a policy threshold at size $\bar{S}$, which discretely increases costs to those above the threshold. This could be for example a regulatory threshold, where firms above the threshold are subject to additional regulation, or it could be a welfare cliff, abruptly ending benefits for individuals. For now, we simply assume that this policy notch results in a lower growth rate for units in a small interval below the threshold $[\bar{S} - \delta, \bar{S}]$. The growth rates below and above this interval are assumed to be unchanged. We discuss potential micro-foundations for this behavior in Section 2.1. This results in drift terms of the form:

$$g_1(S_t) = \begin{cases} g(S_t), & \text{if } S_t \notin [\bar{S} - \delta, \bar{S}] \\ g(S_t) - \gamma, & \text{if } S_t \in [\bar{S} - \delta, \bar{S}]. \end{cases} \tag{2}$$

The introduction of such a threshold may induce strategic behavior among units, discouraging growth beyond the policy limit. Such behavior is likely to occur in practice, as

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

units may seek to avoid crossing the regulatory threshold under certain circumstances, for instance, due to the fixed costs it entails. This incentive is particularly salient in our empirical application context, where exceeding the payroll tax threshold can impose substantial fixed costs on firms that are tied to underlying labor expenses.

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

In Appendix B, we derive the result that for an Itô diffusion in Equation (1) with a growth barrier as in Equation (2), the counterfactual distribution in logs is:

$$\log f_0(S) = \begin{cases} \log(1-p) - \log(1-p(1-F_1^+)) + \log f_1(S), & \text{if } S \in [S_{\min}, \bar{S}-\delta[\\ \log(1-p) - \log(1-p(1-F_1^+)) - \frac{2\gamma}{\sigma^2}[S-(\bar{S}-\delta)] + \log f_1(S), & \text{if } S \in [\bar{S}-\delta, \bar{S}] \\ -\log(1-p(1-F_1^+)) + \log f_1(S), & \text{if } S \in]\bar{S}, \infty]. \end{cases} \tag{3}$$

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 $F_1^+ = \int_{\bar{s}}^{\infty} f_1(s)ds$ is the observed share of units above the threshold.

Hence, when δ is small, an estimate of $\log f_0(S)$ can be obtained using estimates of p and the moments of the observed firm distribution $f_1(S)$:

$$\hat{\log} f_0(S) = \begin{cases} \log(1-\hat{p}) - \log\left(1-\hat{p}(1-F_1^+)\right) + \log f_1(S), & \text{if } S < \bar{S} \\ -\log\left(1-\hat{p}(1-F_1^+)\right) + \log f_1(S), & \text{if } S > \bar{S}. \end{cases} \tag{4}$$

The theoretical result in Equation (3) describes the shifts in the distribution caused by a threshold that acts as a growth barrier. These shifts take a particularly simple form when only units close to the threshold are affected (i.e. when δ is small) and can then be expressed in terms of two estimable parameters: the probability p that a unit which otherwise have grown beyond the policy threshold does not do so, and the share of units above the threshold, F_1^+, when the threshold is in place. Using these parameters, the observed distribution $f_1(S)$ can be adjusted to construct a counterfactual distribution $\hat{f}_0(S)$ that would prevail in the absence of the threshold as in Equation (4).

The intuition for the result is that when a growth barrier is introduced at $\bar{S}$, fewer units grow beyond the threshold. Consequently, the number of units above the threshold declines,

while the number of units below it increases relative to the counterfactual distribution.[2] Importantly, as Equation (4) implies, if we observe a distribution $f_1(S)$ which is distorted by a growth barrier, and we have estimates of p and F_1^+, we can construct an estimate of the counterfactual distribution $f_0(S)$.

As p and F_0^+ are between 0 and 1, the threshold shifts the distribution below the threshold up, while the distribution above the threshold gets shifted down. Moreover, the level difference between these shifts in loq pdfs is $\log(1-p)$.

Figure 1 illustrates these shifts in the context of a firm distribution. The figure shows the observed and counterfactual log pdfs by the log size of firms, where the low-growth area is illustrated by a shaded red area between $\bar{S}-\delta$ and the policy threshold at $\bar{S}$. The observed share of firms above the threshold is illustrated as the blue shaded mass below the observed distribution $f_1(S)$ (solid black line). The counterfactual distribution $f_0(S)$ is illustrated by the dashed black line.

2.1 Microfoundations

The theoretical framework we apply is completely mechanical: the framework does not rationalize the behavior of firms. We simply note that if firms behave in this way – i.e. some fraction p of firms that would have grown past the threshold do not do so due to the regulation – then the firm distribution changes in a simple and predictable way. This obviously leaves the question of why firms would behave in this way. In this subsection, we provide a simple way of rationalizing such behavior: reduced incentives for searching for new business opportunities just below the threshold. The example is kept simple by design, so as to provide the intuition behind it.

Reduced incentives to search for business opportunities. We mostly follow the two-period setup in Aghion et al. (2023b), although we abstract away from many of the details to focus on the decision of growing past the threshold. Aghion et al. (2023b) relate their model to innovation, but the model studied here could be interpreted as a search model, where the firms grow by searching for some general business opportunities. This could be interpreted e.g as search for new employees, customers, investment opportunities, ideas etc.

Assume that firms live for two periods, and use input n to produce revenue $R(n)$.[3] We assume that the input is discrete, and that the firm can search in the first period by choosing a costly search intensity z. Search is successful at probability z, and leads to the firm operating

[2]Thus far, our analysis focuses exclusively on intensive margin responses and growth effects, omitting extensive margin responses, such as firm entry and exit.

[3]In Aghion et al. (2023b), n is the number of product lines in which the firm operates in and labor inputs are proportional to n. Here we simply concentrate on some productive input n.

at a larger scale $(n+1)$ in the second period. However, with probability x the firm may also exogenously end up operating at a lower scale $(n-1)$. The firm incurs a cost $C(n,z)$ from the use of the inputs and the search activity, such that the cost function is increasing and convex in both n and z.

The firm owner then maximizes the expected net present value of the firm over its search intensity $z \geq 0$:

$$\max_{z \geq 0} \{R(n) - C(n,z) + \beta E[\pi(n')]\}, \tag{5}$$

where β is the discount factor of the firm, n' is the second period input usage of the firm, and $\pi(n') = R(n') - C(n',0)$ is the second period profit of the firm.

Using the probabilities of changing scale, this can be written as

$$\max_{z \geq 0} \{R(n) - C(n,z) + z\beta[\pi(n+1) - \pi(n)] + x\beta[\pi(n-1) - \pi(n)] + \beta\pi(n)\}. \tag{6}$$

Here, $z\beta[\pi(n+1) - \pi(n)]$ is the probability of a successful search times the incremental increase in the present value of profits that it provides, and $x\beta[\pi(n-1) - \pi(n)]$ is the probability of exogenous downsizing times the incremental decrease in the present value of profits that it induces.

Assuming that the FOC of this problem describes a global maximum, we can characterize the optimal level of search activity at input level n by

$$\frac{\partial C(n,z^*)}{\partial z} = \beta[\pi(n+1) - \pi(n)], \tag{7}$$

which states that at the optimal level of search (z^*) the marginal cost of search activity should be equal to the present value of the increase in profits of a successful search.

What then happens if the government decides to impose a fixed cost F for firms above a given size $n > \bar{n}$? In this case, only the decision environment of firms operating just below the threshold at $\bar{n}$ changes. They will have the following FOC:

$$\frac{\partial C(\bar{n},z)}{\partial z} = \beta[\pi(\bar{n}+1) - F - \pi(\bar{n})], \tag{8}$$

where the only change is that now the future profits of a successful search are reduced by the fixed cost F. Now, as the cost function is convex (so that $\partial^2 C(\bar{n},z)/\partial^2 z > 0$), the introduction of the fixed cost F will reduce search activity at $\bar{n}$, but not affect it at other levels of input $n \neq \bar{n}$.[4]

[4]To see this, note that F does not enter the FOC of firms operating above $\bar{n}$, as F is paid both in the present state of such firms $(n > \bar{n})$ and in the state they would grow to $(n+1 > n > \bar{n})$. Nor does it enter the FOC of firms operating further below the threshold $(n < \bar{n})$ as they would only grow to $\bar{n}$ at most, and

Denote the choice of search activity at input level n as a function of the government imposed fixed costs by $z^*(n, F)$. We then have that in this model p – the probability that a firm which otherwise would have grown past the threshold does not do so due to the threshold – is given by

$$p = \frac{z^*(\bar{n}, 0) - z^*(\bar{n}, F)}{z^*(\bar{n}, 0)}, \tag{9}$$

thus providing one way to rationalize the behavior of firms in our mechanical framework.

3 Difference-in-differences methodology

The theoretical framework in Section 2 demonstrates that policy thresholds can act as growth barriers, reshaping entire distributions rather than just creating local bunching. This insight motivates our empirical approach but also creates methodological challenges. When a threshold is removed, units may relocate within the distribution or change their survival decisions, making it difficult to identify which units were "treated" by the policy. This section develops a methodology to address these identification challenges.

When discrete policy thresholds induce units to relocate within a distribution, either crossing thresholds they previously avoided or exiting due to the threshold, standard quasi-experimental research designs face an identification challenge. In such settings, it becomes difficult to precisely define which units were effectively treated by the policy change. The identification challenge arises because the policy change affects both the location and existence of units within the distribution. When a payroll tax notch affecting firms is repealed, for instance, some firms may cross the threshold they previously avoided, while others may remain in the market rather than exit. These "switcher" units were effectively treated by the reform, yet standard difference-in-differences strategies would misclassify them as untreated in the pre-reform period. This contaminates the control group and leads to biased estimates.

We propose a method to estimate bounds on the impacts of such reforms under alternative assumptions about how switching is affected. The approach is general but we illustrate it using our empirical application: the repeal of a firm-level payroll tax notch in Finland. In Section 6, we apply this methodology to estimate the effects of this reform on firm scale, input demand, and incidence measures using detailed administrative data.

We begin by illustrating the identification challenge (Section 3.1), then define our causal parameter of interest (Section 3.2), establish the necessary assumptions (Section 3.3), and finally present estimators for two polar cases that provide bounds on the treatment effect (Sections 3.4–3.5).

would not pay the fixed cost in any case.

3.1 An identification issue

When a policy notch is removed or put into place, which units are treated? As units move along the distribution, the answer to this question is not self-evident and has important implications for identifying the effects of such reforms. Standard empirical strategies such as difference-in-differences-type methods and instrumental variable methods typically 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 notch would suffice to determine whether it was treated (or would have been treated in pre-reform periods had the reform already been implemented). However, when notched policy schedules are removed, they may induce units to cross thresholds they previously avoided and reduce exit rates among units above the threshold. As we document empirically in Section 5, these patterns are precisely what we observe in our data. These switcher units are effectively treated by a reform, as the policy change directly induces them to alter their behavior. Nonetheless, standard identification methods classify the threshold-crossing units as untreated prior to the reform since they are located below the notch and would not have classified the units that exited because of the threshold at all.

To address this issue, we define the treatment group of units in a given period to include all units that in the absence of the policy threshold would have located above the threshold. This group includes both units that stay above the notch regardless of whether the notch is in place, which we call *above-stayers*, and two types of switchers: units that have refrained from crossing the notch, which we call *notch-switchers*, and units that exit due to the notch, which we call *exit-switchers*. The stylized movements of these groups in the units in the context of firm distributions are illustrated in Figure 2.

We cannot directly observe which individual units are switchers, since we do not observe where each unit would have been located in the distribution in the absence of the notch (or, conversely, with the notch in place during the pre-reform period). Consequently, we cannot classify units as treated or untreated in a straightforward manner, as would be required in a standard difference-in-differences framework.[5]

To overcome the issue that the true treatment status cannot be directly observed, we propose a method to estimate the effects of policy notch removals on treated units. The key idea is that, even without directly observing treatment status, changes over time in mean outcomes and units counts above and below the threshold are informative about the influence of switchers under certain assumptions. We consider two polar cases: one in

[5]This same limitation applies to Gruber-Saez-style instruments (Gruber and Saez, 2002), as a conventional IV approach also requires observing the actual treatment status of each firm in order to construct a valid instrument for it.

which all switchers are exit-switchers and another in which all are notch-switchers. These assumptions yield lower and upper bounds for the case in which both types coexist. Figure 3 illustrates the stylized post-reform and counterfactual distributions corresponding to each case in the context of firms.

First, consider the case in which all switchers are exit-switchers. Panel (a) of Figure 3 provides a stylized illustration of this scenario. In this case, the distribution below the notch is identical both with and without the notch, that is, the post-reform and counterfactual distributions coincide below the threshold. Units below the notch therefore provide a clean control group. However, what standard methods fail to account for is that the removal of the notch also increases the number of units above the threshold. If, for instance, the mean outcomes of the exit-switcher units were the same as those of the above-stayers, a standard repeated cross-sections difference-in-differences estimator would yield a treatment effect on the treated of zero. This would be incorrect, since some units that actually exist post-reform would have exited had the notch remained in place. For many outcome variables, the counterfactual value for an exit-switcher would thus be zero, while the observed value is non-zero, implying a non-zero treatment effect.

Our method accounts for this by estimating the number of exit-switchers from the distribution, using an appropriately adjusted pre-reform distribution as the counterfactual. We then compute the average treatment effect on the exit-switchers by assuming that their counterfactual outcome is zero, while their observed mean outcome equals that of the above-stayers.

The second scenario, in which all switchers are notch-switchers, is illustrated in Panel (b) of Figure 3. Under this assumption, the total number of units is the same in the observed and counterfactual distributions. This allows us to construct a counterfactual firm distribution by appropriately adjusting the pre-reform distribution. By comparing the actual post-reform distribution to this counterfactual, we can infer the number of notch-switchers. Assuming that the outcomes of the notch-switchers are equal to those of the below-stayers when they are located below the notch then enables estimation of the overall treatment effect.

This assumption could be violated if there were substantial bunching near the threshold, implying that units just below the notch differ systematically from other units below it. In our application in Section 5 we do observe some bunching, but its magnitude is limited relative to the large changes observed in the firm distribution above the threshold before and after the reform.

3.2 Causal parameter

Our causal parameter of interest is the logarithmic average treatment effect on the treated, or $\log ATET$, which is defined as

$$\log ATET = \log(E(y_f(1)|f \text{ is treated})) - \log(E(y_f(0)|f \text{ is treated})), \tag{10}$$

where $y_f(R)$ represents the potential outcome y for unit f under reform status $R \in \{0, 1\}$, with $R = 1$ denoting exposure to the reform. Thus, $\log ATET$ captures the log difference in expected potential outcomes for treated units with and without the reform.

The reason we define our parameter of interest in this way is twofold. First, empirically, the parallel trends assumptions we rely on hold for pre-trends in log outcomes. However, since some units may not operate in the absence of the reform (i.e., when $R = 0$), their potential outcomes $y_f(0)$ would be zero for many variables. Taking logs at the unit level would then imply potential outcomes tending toward $-\infty$. By taking logs at the aggregate level, as in Equation (10), we avoid this issue in most cases.[6]

Second, $\log ATET$ has an appealing interpretation: it approximates the aggregate percentage effect of the reform on the treated. This contrasts with estimating a standard unweighted ATET using unit-level log-outcomes, which could diverge from the aggregate impact if, for example, smaller and larger units responded differently to the reform.

Under two distinct sets of assumptions, we show that $\log ATET$ is point-identified. The difference between these assumptions lies in whether all switcher units are assumed to have been induced to cross the notch by the reform, or instead to have remained in existence (i.e., not exited) because of the reform. In cases where both types of switching occur, these two estimates define the lower and upper bounds for $\log ATET$.

3.3 Setup

To simplify the exposition we consider a case with two periods t, a pre-reform period $t = 0$ and a post-reform period $t = 1$. The realized (and observed) outcome of unit f in period t is given by $y_{f,t}$. In each period, we assume that there are four groups of units: 1) units that would stay below the notch point regardless of whether the notch would be in place in that period, or *below-stayers*; 2) units that would stay above the notch regardless of the reform, or *above-stayers*; 3) units that are below the notch if the notch is in place in this

[6]This issue has been recently analyzed in detail by Chen and Roth (2024), who propose three alternative solutions. Estimating $\log ATET$ provides another practical remedy, although it does not entirely eliminate this issue: if the average potential outcome among the treated were zero, the problem of logging zero would re-emerge. In empirical applications, however, this situation is rare.

period, but are above the notch if the notch is not in place in this period, or *notch-switchers*, and 4) switchers that are in non-existence if the notch is in place in this period, but are above the notch if the notch is not in place in this period, or *exit-switchers*. These groups are illustrated in Figure 2, which compares hypothetical distributions of firms with (solid black line) and without (dashed black line) the notched schedule. The difference between these arises from the notch-switchers and the exit-switchers. Removing the notch induces notch-switchers to relocate above the notch (from the light red to the red area) and the exit-switchers to not to exit (from non-existence to the blue area). Below-stayers (light yellow area) and above-stayers (light green area) stay where they are located.

Group identifiers for these different types of units are $g \in \{bs, as, ns, es\}$ respectively. The number of units in group g in period t is n_t^g ($\equiv \sum_f 1(f \in g \text{ during } t)$), and the average outcome of the group is y_t^g ($\equiv E(y_{f,t}|f \in g \text{ during } t)$). Switchers and above-stayers are defined as treated by the reform, as they would have located above the notch in the absence of the reform. Below-stayers are non-treated. In a standard difference-in-differences setup, we would only have below- and above-stayers. This does not mean that a standard difference-in-differences approach would not allow units to move in the distribution. What these methods rule out is that the reform changes this movement across the notch.

We use potential outcome notation, so that $y_t^g(R) = E(y_f(R)|f \in g \text{ during } t)$ 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 ($R = 1$). However, note that R is not a treatment status indicator, but an indicator of whether or not the reform is in place. The actual treatment group consists of three groups of units: the above-stayers, the notch-switchers and the exit-switchers. This means that $\log ATET$ in the post-reform period $t = 1$ can be expressed as

$$\begin{aligned} \log ATET &= \log(E(y_{f,1}(1)|f \text{ is treated})) - \log(E(y_{f,1}(0)|f \text{ is treated})) \\ &= \log(E(y_1^g(1)|g \in \{as, ns, es\})) - \log(E(y_1^g(0)|g \in \{as, ns, es\})). \end{aligned} \tag{11}$$

To identify $\log ATET$, we need to make the following assumptions:

Assumption 1 (Parallel trends in log numbers of firms).

$$\log(n_1^{bs}) - \log(n_0^{bs}) = \log(n_1^{as}) - \log(n_0^{as}) = \log(n_1^{ns}) - \log(n_0^{ns}) = \log(n_1^{es}) - \log(n_0^{es}) \tag{12}$$

Intuitively, this assumption states that the number of units in each group would grow with the same growth rate, or that the share of each group stays the same. Note that for exit-switchers this includes also units that do not exist, but would exist if the reform would not be in place. We can use data to examine the validity of this assumption by examining

how the log number of units above and below the threshold developed pre-reform. The assumption implies that the log number of units below and above the threshold develops similarly without the reform. However, similarly as in standard parallel trends assumptions, the assumption is fundamentally untestable as it is an assumption on unobservables: similar pre-trends are only suggestive evidence that the assumption might hold.

Assumption 2 (Parallel trends in log mean outcomes).

$$\log(y_1^{bs}(0)) - \log(y_0^{bs}(0)) = \log(y_1^{as}(0)) - \log(y_0^{as}(0)) = \log(y_1^{ns}(0)) - \log(y_0^{ns}(0)) \tag{13}$$

This assumption is similar to the standard parallel trends assumption in means although it needs to apply for all the different groups separately except for the exit-switchers. In other words, the assumption is that the log mean outcomes of each of the three groups (below- and above switchers and notch-switchers) would have developed similarly had the reform not taken place. As without the reform notch-switchers stay below the notch, this implies that we can examine the validity of this assumption in the form of parallel pre-trends.

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

$$\begin{cases} y_t^{ns}(0) = y_t^{bs} \\ y_t^{es}(0) = 0 \end{cases} \tag{14}$$

This assumption has two parts. The first part states that the mean outcome of notch-switchers matches that of below-stayers when the notch-switchers are below the notch. The second part states that the mean outcome of exit-switchers is zero when the notch is in place ($R = 0$) and they are not in business. One way to provide suggestive evidence to support the first part of this assumption is to look at whether the outcome-distribution changes substantially below the notch post-reform. If, for example, the log distribution and log means by bin change similarly across the distribution below the notch, this suggests that notch-switchers could have had a similar distribution of outcomes below the notch as the below-stayers, and hence a similar mean outcome.

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

$$y_t^{bs}(0) = y_t^{bs}(1) \tag{15}$$

This assumption states that the mean outcomes of the below-stayers are not affected by the reform. This is a variant of the standard stable unit treatment value assumption (SUTVA), although slightly less strict. This is less strict because we do not need to assume that treated units are not affected by the treatment status of other treated units as in the

standard SUTVA. In other words, we do not need to assume that there are no spillovers between the treated units. This is because we are interested in the effect of the reform on the treated on the whole, also including any spillovers form one treated to another.

We show that under assumptions 1–4, $\log ATET$ can be expressed as

$$\log ATET = \underbrace{\tau^{DID}}_{\text{Diff-in-diff estimate}} \underbrace{- \log\left(\frac{n_1^{as}}{n_{1,1}} + \frac{y_{0,0}}{y_{1,0}}\frac{n_1^{ns}}{n_{1,1}}\right)}_{\text{Switcher term}}, \tag{16}$$

where $\tau^{DID} = \log(y_{1,1}) - \log(y_{1,0}) - (\log(y_{0,1}) - \log(y_{0,0}))$ is a difference-in-differences estimate comparing log means post- and pre-reform below and above the notch point[7] and $n_{1,1}$ is the observed number of units above the threshold in period $t = 1$. Appendix C provides the detailed proofs. We use the notation $n_{a,t}$ to express the observed number of units above ($a = 1$) or below ($a = 0$) the threshold in period t. Equation (16) is revealing of where the bias in the more standard difference-in-differences estimate τ^{DID} comes from: if there are no switchers then $n_1^{as} = n_{1,1}$ and $n_1^{ns} = 0$ so that the switcher term is zero and the difference-in-differences estimate is unbiased. However, when there are switchers, the switcher term is in general non-zero, which seems evident in our empirical example as described in Section 5.

It is obvious that we do not directly observe n_1^{as} and n_1^{ns} from the data and thus, we need to make further assumptions of the movement of units in response to the reform to be able to estimate $\log ATET$. Therefore, we next turn to two different scenarios, under which we can point-identify $\log ATET$ and which provide bounds for the $\log ATET$ when there are switchers. The first scenario is that all the switchers are exit-switchers and the second is that all the switchers are notch-switchers.

3.4 $\log ATET$ with only exit-switchers

We show that $\log ATET$ is point-identified under the assumption that all switchers are exit-switchers and there are no notch-switchers. Changes in the firm distribution in this case are characterized in Figure 3 panel a. Note that in this case, the reform has no impact on units below the notch, so that the distribution (see Figure A4d) and also the mean outcomes below the notch are unaffected. More formally, in this case, we assume:

Assumption 5 (No notch-switchers.).

$$n_t^{ns} = 0 \tag{17}$$

[7]Note that this corresponds to a weighted cross-sectional difference-in-differences estimate of unit log outcomes if no unit has outcome 0. As we consider cases with exit/entry, some units would have log(0) outcomes at times, which would complicate the analysis.

Under assumption 5, the expression for $\log ATET$ in Equation (16) becomes (see Appendix C for derivation):

$$\log ATET = \tau^{DID} + \tau_n^{DID} = \tau_{agg}^{DID}, \tag{18}$$

where $\tau_n^{DID} = \log(n_{1,1}) - \log(n_{1,0}) - (\log(n_{0,1}) - \log(n_{0,0}))$ is the difference-in-differences estimate of log number of units comparing pre- and post-reform periods above and below the notch, and $\tau_{agg}^{DID} = \log(Y_{1,1}) - \log(Y_{1,0}) - (\log(Y_{0,1}) - \log(Y_{0,0}))$ is the analogous difference-in-differences estimate for aggregate outcomes ($Y_{a,t} = \sum_{f \in a} y_{ft}$).

The intuition in this case is that when there are only exit-switchers, everything we observe below the notch just reflects changes among the below-stayers which are not affected by the reform (by assumption 4). Hence, the behavior of units below the threshold provide a counterfactual for changes above the threshold when the parallel trends assumptions 1 and 2, and the SUTVA-style assumption 4 hold. The assumption that the outcome of exit-switchers would have been zero in the absence of the reform (from assumption 3) then allows us to calculate the average change among the treated (exit-switchers and above-stayers) and compare this to the counterfactual changes below the threshold.

3.5 $\log ATET$ with only notch-switchers

We derive another bound for $\log ATET$ under the assumption that all switchers are notch-switchers, or that there are no exit-switchers. The corresponding assumption is:

Assumption 6 (No exit-switchers.).

$$n_t^{es} = 0 \tag{19}$$

Under assumption 6, the expression for $\log ATET$ in Equation (16) becomes (see Appendix C for derivation in detail):

$$\log ATET = \tau^{DID} + \tau_n^{DID} + \Psi, \tag{20}$$

where Ψ is a correction term equal to $\log\left(\frac{f_{0,1}}{f_{0,0}}\right) - \log\left(1 + \frac{y_{0,0}}{y_{1,0}}\left(\frac{f_{1,1}}{f_{1,0}} - 1\right)\right)$. Here $f_{a,t}$ is the share of units above ($a = 1$) or below ($a = 0$) the threshold in period t.

The intuition behind this result is that if the switchers are all notch-switchers, we need to adjust Equation (18) for two reasons: first, the potential outcome of the notch-switchers when the reform is not in place might not be zero as with the exit-switchers. Second, as illustrated in Figure 3, the unit distribution changes not only above the notch, but also below

the notch. This is because notch-switchers are moving from below the notch to above the notch, hence reducing the number of units below the notch and increasing the number of units above. This means that if we would simply take the distribution below the notch as a valid counterfactual (as we do with the exit-switchers in panel a), we would overestimate the change in the number of units above the notch and underestimate the change below the notch. The correction term Ψ adjusts the estimate to take these two channels into account.

4 Institutional Background and Data

Having presented the theoretical framework and estimation methodology in Sections 2–3, we now provide details on our empirical application: the Finnish payroll tax system, which featured a size-based threshold that created a tax notch for firms above a certain cutoff. This section describes the institutional setting and data used to implement the framework developed above.

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 euros, above which the higher rate applied.

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 3 percentage point increase in their payroll tax rate. Appendix A, Table A1 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 very small businesses whose payroll tax rates remained unchanged.[8]

In January 2010, the threshold was repealed: after the reform, all firms, regardless of whether their capital depreciation was below or above 50,500 euros, 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.

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 plant-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

[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 euros 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.

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, etc. These data are available from 1987 onward 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 99th or below the 1st percentile of each examined continuous outcome variable. 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 5,000 and 450,000 euros of annual depreciations and for the last 20% of firms just below the threshold in 2005–2007.

Finally, Figure 4 shows the variation we exploit in our analysis described in detail in Section 4.1, including all firm-level payroll taxes, in percentage points across the firm-level capital depreciation distribution. 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 a 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 across the entire distribution.

5 Testing the Theory: Firm Distributions and Dynamics

This section tests the theoretical predictions from Section 2 and provides empirical support for the assumptions underlying our estimation methodology from Section 3. We apply the theoretical framework to examine whether the payroll tax threshold functioned as a growth barrier. If it did, we should observe two key patterns. First, the threshold should constrain

firm growth by reducing the probability that a firm crosses it. We study this prediction using the dynamic bunching approach developed by Garbinti et al. (2023), which allows us to estimate the probability p that a firm which would otherwise have crossed the threshold refrains from doing so. 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.

5.1 Estimating the sufficient statistic p with dynamic bunching

To estimate p, we define the treatment group as the three log depreciation bins closest to, but below, the threshold. This group is selected because it exhibits significant bunching in growth rates near the threshold. The control groups consists of the next three bins further below the threshold. Depreciation values in the treatment group range from €22,715 to €50,499, while those in the control group range from €15,210 to €22,714.

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

$$\bar{g}^1_{i,t} = \frac{d_{i,t+1} - d_{i,t}}{d_{i,t}} - \frac{50500 - d_{i,t}}{d_{i,t}} = \frac{d_{i,t+1} - 50500}{d_{i,t}}, \tag{21}$$

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 in order to construct a counterfactual distribution for the treatment group. More specifically, we define the counterfactual normalized growth rates of the control group as

$$\bar{g}^0_{i,t}(\hat{\alpha}) = \frac{d_{i,t+1} - \hat{\alpha} \times 50500}{d_{i,t}}, \tag{22}$$

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^*. 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, our results are robust to alternative choices of this parameter. In

other words, $\hat{\alpha}$ is chosen by

$$\hat{\alpha} \equiv \arg\max_{\alpha} \frac{1}{N_b} \sum_{b \in [b_{\min}, g^*]} [H_1(b) - H_1^0(b|\alpha)]^2, \tag{23}$$

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 $H_1^0(b|\alpha)$ are the empirical CDF's of normalized growth rates $\bar{g}_{i,t}^1$ and $\bar{g}_{i,t}^0(\alpha)$.

Figure 5 shows the distributions of $\bar{g}_{i,t}^1$ for the treatment group and $\bar{g}_{i,t}^0(\hat{\alpha} = 0.535)$[9] for the control group. Panel 5a shows the PDF's of these distributions and 5b shows the corresponding CDF's. 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

$$\hat{p} = \frac{\text{Bunching mass}}{\text{Counterfactual mass above threshold}} = \frac{H_1(0) - H_1^0(0|\hat{\alpha} = 0.535)}{1 - H_1^0(0|\hat{\alpha} = 0.535)}. \tag{24}$$

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 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.

5.2 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 6 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).[10] The bin size is defined such that

[9] Appendix Figure A1 shows the mean square errors (MSE) from Equation (23) for different values of the scaling parameter α. The MSE reaches its minimum at a scaling parameter of approximately 0.535, which we use to construct the counterfactual normalized growth distribution. The results are robust for the scaling parameter with different choices of g^* and with minimizing the MSE between PDF's instead of CDF's.

[10] 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

each bin corresponds to 20 log points of capital depreciation, and this same bin definition is used consistently throughout the paper.

The figure shows that, during 2005–2007, the firm-size distribution exhibits pronounced 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 remarkably well with the observed distribution in the post-reform period 2011–2013. It is important to emphasize that 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 relied 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 adjusted 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 6 also illustrates that firm distributions by depreciation follow a power law[11], 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 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.

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 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

organizations were involved in drafting the legislation, some firms may have anticipated the reform as early as 2008.

[11] 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.

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.[12]

5.3 Are These Responses Real?

A potential explanation for these firm responses 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 7 shows a very similar firm distribution when using depreciation reported in accounting data, as in Figure **??**. This accounting-based measure is less prone to manipulation that 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) Figure 7, 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 of bins (as illustrated in Panel (d) of Figure 7). 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 7, 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 A3 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 A3, which plots

[12] Appendix Figure A4 further investigates firm movement across the capital depreciation distribution before and after the payroll tax discontinuity. Panel (a) of Figure A4 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 A4 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.

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 A3, 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.

Summary. 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.[13] 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 learning-by-doing dynamics, where firms must first learn to operate at a slightly smaller scale before expanding further, a notch could also lead to fewer firms above the threshold even in the absence of extensive-margin effects.

Having documented this pattern of firm movement and validated our theoretical predictions, we now quantify the causal effects of removing the notch. The next section applies the bounding methodology from Section 3 to estimate treatment effects on firm scale, input

[13]Moreover, the employment responses documented in Section 6.2 and Section 6.3 below indicate that the observed effects of the payroll tax increase are real economic responses rather than reporting or avoidance behaviors.

composition, and tax incidence.

6 Empirical Results: Firm Scale, Input Demand and Incidence Responses

Having established in Section 5 that the payroll tax notch acted as a growth barrier with substantial firm movement across the threshold, we now apply the estimation methodology developed in Section 3 to quantify the causal effects of the reform. This section first discusses the implementation of that methodology in our empirical setting (6.1, it then presents results on firm scale responses (Section 6.2), compositional changes in inputs (Section 6.3), and tax incidence (Section 6.4).

6.1 Implementation of the difference-in-differences methodology

We illustrate the impacts of the reform by plotting bin-level estimates of $\log ATET$ across bins of capital depreciation first with the exit-switchers only. We use the same 20 log point bins of capital depreciation as we do in Section 5. This provides a way to graphically examine how the reform affected outcomes across the distribution. More formally, we compare log aggregate outcomes in bin b to bin c:

$$\widehat{\log ATET_b}^{es} = \Delta \log \sum_{f \in b} y_f - \Delta \log \sum_{f \in c} y_f. \tag{25}$$

This can be decomposed into two parts that can be plotted separately: a change in the log number of firms and a change in the log mean as

$$\widehat{\log ATET_b}^{es} = \Delta \log \frac{1}{n_b} \sum_{f \in b} y_f - \Delta \log \frac{1}{n_c} \sum_{f \in c} y_f + \Delta \log n_b - \Delta \log n_c. \tag{26}$$

These provide valid bin-level estimates if there are only exit-switchers. To get a more comprehensive view of the impacts of the reform, we also provide point-estimates of the full $\log ATET$ comparing pre- and post-reform outcomes below and above the threshold both when there are only exit-switchers and when there are only notch-switchers. The corresponding estimators are:

$$\begin{aligned} \widehat{\log ATET}^{es} &= \Delta \log \sum_{f \in \text{above}} y_f - \Delta \log \sum_{f \in \text{below}} y_f \\ &= \widehat{\tau^{DID}} + \widehat{\tau_n^{DID}}, \end{aligned} \tag{27}$$

and

$$\begin{aligned}\widehat{\log ATET}^{ns} &= \Delta \log \sum_{f \in \text{above}} y_f - \Delta \log \sum_{f \in \text{below}} y_f + \hat{\Psi} \\ &= \widehat{\tau^{DID}} + \widehat{\tau_n^{DID}} + \hat{\Psi},\end{aligned} \tag{28}$$

where above refers to bins above the threshold and below refers to bins below the threshold. $\widehat{\tau^{DID}}$, $\widehat{\tau_n^{DID}}$, and $\hat{\Psi}$ are the sample equivalents of τ^{DID}, τ_n^{DID}, and Ψ in Equation (20). We exclude the closest bin to the threshold both from those above and those below to mitigate the impact that bunching might have on the estimate.

Standard errors are estimated using a bootstrap procedure. We resample observations from the data with replacement and create 500 different resampled datasets. We then calculate the estimates for each of these resamples separately and calculate the standard deviation of the estimates. This gives us an estimate of the standard error of the estimate.

As a robustness check, we also include another control group in the form of firms that were not affected by the reform (as described in Section 4, the notch did not affect firms with capital depreciation less than 10% of labor costs). With this additional control group, the estimators become analogous to the triple-difference (DDD) estimator:

$$\begin{aligned}\widehat{\log ATET}^{es} &= \Delta \log \sum_{f \in \text{above}, f \in T} y_f - \Delta \log \sum_{f \in \text{below}, f \in T} y_f \\ &\quad - \left[\Delta \log \sum_{f \in \text{above}, f \in C} y_f - \Delta \log \sum_{f \in \text{below}, f \in C} y_f \right] \\ &= \Delta \widehat{\tau^{DID}} + \Delta \widehat{\tau_n^{DID}},\end{aligned} \tag{29}$$

and

$$\begin{aligned}\widehat{\log ATET}^{ns} &= \Delta \log \sum_{f \in \text{above}, f \in T} y_f - \Delta \log \sum_{f \in \text{below}, f \in T} y_f \\ &\quad - \left[\Delta \log \sum_{f \in \text{above}, f \in C} y_f - \Delta \log \sum_{f \in \text{below}, f \in C} y_f \right] + \hat{\Psi}_{f \in T} - \hat{\Psi}_{f \in C} \\ &= \Delta \widehat{\tau^{DID}} + \Delta \widehat{\tau_n^{DID}} + \Delta \hat{\Psi},\end{aligned} \tag{30}$$

where $\Delta\widehat{\tau^{DID}}$, $\Delta\widehat{\tau_n^{DID}}$, and $\Delta\hat{\Psi}$ are the differences of $\widehat{\tau^{DID}}$, $\widehat{\tau_n^{DID}}$, and $\hat{\Psi}$ of the firms that were affected by the reform compared to the firms that were unaffected by the reform. This additional control group allows us to better control for any macroeconomic trends that might, for example, impact larger firms more than smaller firms.

6.2 Wages, Employment and Firm Scale

As is standard in the payroll tax incidence literature, we first examine the average net-of-payroll tax earnings of employees, capturing the effect of payroll taxes on worker-level wages across depreciation distribution bins before and after the reform. Panel (a) of Figure 8 shows an overall upward trend in wages across the depreciation distribution before and after the repeal of the notch. However, there is no clear discontinuity in average wages at the threshold – neither when the notch was in effect nor after its removal. Panel (c) shows a slight increase in wages among firms above the threshold when comparing pre- and post-reform levels, suggesting that firms, at most, modestly increased worker-level wages following the reduction in payroll tax burden.

Two important caveat are worth noting in this analysis. First, 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 – a question we address in greater details in Section 6.3. Second, we also observe a difference in average wages before and after the reform among very small firms – those with depreciation below €20,000 – which is larger than the corresponding difference in wages above the threshold, even though these firms were not directly affected by the policy.

Panel (b) of Figure 8 shows a sharp drop in firm-level labor costs at the threshold, where the payroll tax rate increased for firms. The figure also reveals a strong concentration of firms just below the threshold, with these firms exhibiting considerably higher average labor costs compared to nearby bins in the distribution. After the repeal of the threshold, the distribution of labor costs becomes smoothly increasing across the depreciation bins, indicating that firms above the threshold exhibit higher aggregate wage costs per bin following the reform.

We next decompose these aggregate changes into their underlying components, changes in the number of firms and changes in mean outcomes within bins, following the approach outlined in Equation (26). Panel (d) of Figure 8 displays the overall before-after differences across depreciation bins (blue line), as well as their decomposition into frequency effects (green dashed line), cross-sectional DiD estimates (red line), and log mean changes in employee-level labor costs among firms (orange dashed line). The results show that the overall change in labor costs is primarily driven by changes in the number of firms, while average employee-level wage costs above the threshold remain virtually unchanged. This suggests that the observed responses are largely explained by shifts in the firm distribution rather than by wage adjustments.

Panel (d) of Figure 8 also reports the estimated average treatment effects, based on Equations (27) and(28), under the two extreme cases of only exit-switchers and only notch-

switchers. The corresponding estimates are sizable, 0.133 (SE 0.015) and 0.085 (SE 0.013), respectively.

As average employee wages respond only relatively little to the payroll tax decrease while firm-level labor costs change considerably, the adjustment must occur through changes in firm employment. Panels (a) and (c) of Figure 9, structured as in Figure 8, confirm this: the average number of employees drops sharply at the threshold when the notch is in place, but the discontinuity vanishes after its repeal. The estimated average treatment effects, assuming only exit-switchers and only notch-switchers, are 0.103 (SE 0.014) and 0.048 (SE 0.013), respectively. Thus, the reform clearly reshaped the firm distribution and increased firm size through higher employment.

Panels (c) and (d) of Figure 9 show that these scale effects are also evident in firms' capital stocks above the threshold, with estimated treatment effects of 0.151 (SE 0.026) and 0.098 (SE 0.025), respectively. The figures further show that the changes in the number of employees and capital stock below the threshold, i.e. among firms which were unaffected by the reform, are negligible. This supports the validity of our before–after comparison across the firm distribution and suggests that the observed effects are indeed attributable to the payroll tax reform.

Similar scaling effects are apparent also in both annual revenue and value added. As shown in Figure 10, the overall increases in these outcomes are of comparable magnitude, 6–11 log points depending on the assumptions of switcher types.

Table 2 summarizes the point estimates for the different outcome variables. As highlighted above, overall labor costs increase significantly among firms above the threshold, with magnitudes ranging from 8.5 to and 13.3 log points depending on the assumed type of firm movement. This increase in labor costs is mainly driven by an increase in number of employees, although we also find a statistically significant increase in average wages among firm affected by the reform. However, as noted earlier, this increase in average wages may reflect compositional changes in the workforce rather than a direct pass-through of payroll taxes to incumbent workers' wages – a question we explore further in Section 6.3.

This sizeable increase in labor costs and employment are accompanied by comparable growth in other firm size measures. We observe very similar effects for capital stock, revenue and value added, all suggesting that firm size significantly increased in response to the payroll tax cut. Interestingly, however, we find no corresponding changes in taxable annual profits or in dividends paid to firm owners. This indicates that, while firms grew larger, these gains did not translate into higher returns to owners; instead, the payroll tax cut seems to benefit mainly labor through increased employment and slightly higher wages. Section 6.4 approximates the incidence of payroll taxation in more detail by considering also potential

scaling effects due to the reform.

In the lower panel of Table 2, we further show that the estimated effects are qualitatively robust, but slightly larger in absolute values, when restricting the sample to firms closer to the threshold with annual capital depreciation between €12,500 and €125,000.

One potential concern for our empirical approach would be if firms are somehow differently affected by the underlying shocks across the distribution, for example, through heterogeneous responses to macroeconomic conditions or that behavioral responses correlate with firm size. The sharp changes in all baseline outcomes at the threshold alleviate most such concerns. Similarly, the absence of changes below the threshold supports the robustness of our findings, aside from a small but significant decline in the closest bin below the threshold, likely reflecting pre-reform selection behavior.

To further test the robustness of our results in detail, we use the fact that some small firms were never subject to the threshold and thus faced no payroll tax change pre- versus post-2009. Incorporating these firms as an additional control group, we re-estimate our models using a triple-difference specification (Equations (29) and(30)). Appendix Figures A5 and A6 show that the results remain rather robust, and Appendix Table A3 presents systematically slightly larger estimates than the baseline results in Table 2. This implies that our baseline estimates are conservative. It is unlikely that these results driven by macroeconomic shocks, because the unaffected firms operate in the same industries and time periods as treated firms.

6.3 Compositional Changes in Firm-Level Inputs

Next, we turn to a more detailed examination of compositional changes in firm-level inputs. We begin by analyzing labor composition effects of the payroll tax reform, distinguishing between different types of workers. We then examine firm-level responses in aggregate capital and decompose these effects across different categories of capital assets.

Labor. Section 6.2 already provided evidence of rather large employment responses among firms affected by the payroll tax cut. Next we divide these firm-level employment responses by two dimensions: routine vs. non-routine and high- vs. low-education employees. Figure 11 shows these responses. Panel (a) of Figure 11 shows that the number of routine employees among affected firms increases less with a lower bound estimate 0.024 (SE 0.019) in comparison to the similar estimate of number of non-routine workers in Panel (b) (0.098 (SE 0.012)). Therefore, this evidence suggests that the share of non-routine jobs, which are typically more complex and require higher human capital, increase in these firms that were affected by the payroll tax cut. This pattern is supported by the fact that the responses

among lower educated are smaller in Panel (c) in comparison to higher educated in Panel (d) with lower bound estimates 0.037 (SE 0.012) vs. 0.072 (SE 0.021), although the difference between estimates is more moderate than by job task groups.

Table 3 further collects the results by the type of job tasks and employee-level education measures in shares of workers in firms. In the table, we also restrict our sample to only those employees for which we can observe the occupation and education information. The main reason for this restriction is that we cannot observe occupation codes for all baseline sample years, especially in years prior to 2005. However, this sample restriction has a rather minor impact on the overall employment response in column (1) 0.104 (SE 0.015) in comparison to our baseline employment response 0.103 (SE 0.014), see Table 2 column (3). The estimates reported in Table 3 confirm the main message from the figures above, after the payroll tax cut, the share of non-routine and highly educated workers increase in the affected firms. However, there seems to be no apparent heterogeneity in responses by gender as is visible from columns (5) and (6). Note however that the point estimates in columns (1)–(4) of Table 3, are somewhat smaller in absolute terms if we restrict our sample to capture only firms with 12,500 – 125,000 euros of annual capital depreciation instead using the baseline estimation sample.

Capital. Next we investigate how firm-level investments develop over the depreciation distribution of firms, and also separate total investment into three major groups: fixed assets, buildings and research and development (R&D). Figure 12 collects the results. Panel (a) clearly shows an increase in the total investments above the threshold, which is consistent with the positive overall capital stock response presented in Figure 9d. Evidently, this impact is mostly due to the changes in fixed assets (Panel (b)), although there are notable increases in R&D investments also among affected firms. However, we do not observe any increase in investments in buildings; if anything, we see a slight decrease. This suggests a moderate reallocation from these investments toward R&D, which is typically considered more productive. This shows that also the capital intensity of firms is affected by the increase in payroll taxes. In addition to these results, we also found some responses in rents, which we discussed in Section 5.3, that most likely reflect responses in capital composition between owning or renting capital is affected by the payroll taxes (see Appendix A Figure A4). However, the level of rental payments is, on average, such a low share of all capital across the firm distribution in comparison to the level of annual investments that we interpret these responses small in comparison to the responses described in Figure 12.

Table A5 further divides the responses into the shares of investments across the three categories. The share of fixed-asset investments does not change in response to the payroll tax rate cut. However, we observe a statistically significant decrease of approximately 3.4%

in building investments, accompanied by an increase of roughly 2.7% in R&D investments. These shifts in investment composition are sizeable, considering that the baseline means of building and R&D investments are relatively low at 16% and 11%, respectively. These findings align well with the results for total investments and the category-level investment responses in (log) levels shown in Figure 12. The point estimates in the lower panel of Table 4, where we restrict the sample to firms with annual capital depreciation between 12,500 and 125,000 euros, are broadly similar to the baseline estimates, although smaller in absolute magnitude, similarly with the patterns observed in Table 3.

6.4 Incidence

The substantial changes in the firm-size distribution before and after the abolition of the payroll tax notch, together with strong evidence of real firm-scale responses, suggest that both capital and labor are strongly affected by the increased labor costs. However, given the magnitude of these overall scaling effects associated with the higher payroll tax rate, a more precise analysis of tax incidence that accounts for these quantity adjustments is required. We therefore proceed to estimate the incidence of payroll taxes on both workers and firm owners.

To provide a conceptual framework, we draw on recent empirical studies estimating the incidence of value-added taxes (Benzarti and Carloni 2019) and minimum wages (Harasztosi and Lindner 2019); our approach is also closely related to the rent-sharing literature (e.g., Card et al. 2015). We begin by defining firm profit (Π) to be:

$$\Pi = R - M - wL - rK - C,$$

where R denotes revenue that is net-of-tax price times the quantity of goods or services produced; M represents materials required for producing revenue; w average wage rate, L is the number of workers, r is the depreciation percentage on capital K; and C captures all other – both variable and fixed – costs needed for production.

After rearranging the terms so that value added created in production ($VA = R - M - C$) appears on the left-hand side, taking differences, and scaling by pre-reform value added VA_{pre}, the profit equation above can be expressed as follows:

$$\frac{\Delta VA}{VA_{pre}} = \underbrace{\frac{\Delta(\Pi + rK)}{VA_{pre}}}_{\text{Owner share of change in VA}} + \underbrace{\frac{\Delta(wL)}{VA_{pre}}}_{\text{Worker share of change in VA}}, \tag{31}$$

where $(\Pi + rK)$ represents the share of value added accruing to firm owners, and wL denotes the share accruing to workers. For the purposes of this analysis, we assume that value added fully captures the incidence on consumers, implying that input prices and marginal costs remain unaffected by changes in payroll taxes.

The above framework outlined is simple and does not account, for example, for other taxes such as corporate, dividend or value added taxes. However, since our empirical variation exclusively concerns payroll taxes, this simple identity serves as a useful tool for illustrating how the incidence of the tax is shared among different economic agents. Moreover, we abstract from distinctions between owned and rented capital, as well as from potential outsourcing or changes in employment structure. As shown in Section **??**, we do not observe substantial responses in proxies for these types of behavior, suggesting that these simplifying assumptions are unlikely to drastically affect our conclusions.

Equation 31 clearly shows that the aggregate percentage change in value added can be decomposed into two components, providing three directly observable measures in our data that can be estimated and used to approximate the incidence borne by firm owners and workers.

Figure 13 shows the average values of these three outcomes across depreciation bins. Below the threshold, changes in value added—and consequently in the shares accruing to owners and workers—are minimal, apart from the expected selection effect in the last bin below the threshold. Above the threshold, however, value added increases clearly on average. A substantial portion of this increase stems from the workers' share, $\Delta(wL)$, indicating that workers bear a larger share of the adjustment relative to firm owners. Nevertheless, most of this response reflects changes in the number of workers L rather than in average wages w, which is the channel typically emphasized in the traditional public economics literature when assessing the incidence of payroll taxes.

We estimate the components of equation 31 using the same empirical approach as for $\log ATET^{es}$ in Section 3. Specifically, to obtain causal estimates of the reform's impact on each component, for every $X \in \{VA, \Pi + rK, wL\}$ we estimate

$$\widehat{\frac{\Delta X}{VA_{pre}}} = \frac{\Delta \sum_{f \in above} X_f}{\sum_{f \in above} VA_{f,pre}} - \frac{\Delta \sum_{f \in below} X_f}{\sum_{f \in below} VA_{f,pre}}, \tag{32}$$

where *above* and *below* refer to firms above and below the threshold, X_f is the firm-level outcome X, and $VA_{f,pre}$ is the pre-reform value-added of firm f.

Table 5 reports the point estimates. The results indicate that the average increase in value added—scaled by pre-reform revenue—is approximately 9.6% among firms above the threshold, where payroll tax rates decreased by 2.5 percentage points (equivalent to roughly 15% of total firm-level payroll taxes). This rise in value added reflects, on average, a 3.5% increase in profits and a 6.1% increase in labor costs at the firm level. These estimates imply that the incidence of the payroll tax cut is roughly divided 37–63 between firm owners and workers, respectively. Importantly, most of the worker-side gain arises from increased employment rather than from higher wages – contrasting with the traditional approach in the public finance literature, which typically attributes worker incidence to wage adjustments.

7 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 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 firm refrains from crossing the threshold when the notch is in place. This probability can be quasi-experimentally identified using the dynamic bunching approach, 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 or exit the distribution entirely. To address this challenge, we develop estimators that provide bounds on the log average treatment effect on the treated under alternative assumptions about the nature

of movement in the distribution. 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 four key findings. The payroll tax notch acted as a significant growth barrier, reducing the number of firms above the threshold by approximately 20 percent, with effects extending to firms eight times the threshold size. Following the notch's repeal, firm scale increased substantially, with labor costs rising by 8.5–13.3 log points, employment by 4.8–10.3 log points, and capital stock by 9.8–15.1 log points. Compositional effects reveal that affected firms increased their use of non-routine and highly educated workers while shifting investments from buildings toward research and development. Finally, our incidence estimates indicate that workers bear approximately 63 percent of the payroll tax burden and owners 37 percent, with the worker-side incidence operating almost entirely through the employment margin rather than wage adjustments.

These findings 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 results 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 a wide range of 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., C. Antonin, S. Bunel, and X. Jaravel (2023a): "The Effects of Automation on Labor Demand. A Survey of the Recent Literature," PSE-Ecole d'Ã©conomie de Paris (Postprint) halshs-03756498, HAL.

Aghion, P., A. Bergeaud, and J. Van Reenen (2023b): "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.

Axtell, R. L. (2001): "Zipf distribution of U.S. firm sizes," *Science*, 293, 1818–20.

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

Benzarti, Y. (2025): "Tax Incidence Anomalies," *Annual Review of Economics*, 17, 615–634.

Benzarti, Y. and D. Carloni (2019): "Who Really Benefits from Consumption Tax Cuts? Evidence from a Large VAT Reform in France," *American Economic Journal: Economic Policy*, 11, 38–63.

Benzarti, Y. and J. Harju (2021a): "Can payroll tax cuts help firms during recessions?" *Journal of Public Economics*, 200.

——— (2021b): "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.

Bozio, A., T. Breda, J. Grenet, and A. Guillouzouic (2023): "Does Tax-Benefit Linkage Matter for the Incidence of Social Security Contributions?" *Working Paper*.

Card, D., A. R. Cardoso, and P. Kline (2015): " Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women *," *The Quarterly Journal of Economics*, 131, 633–686.

Chen, J. and J. Roth (2024): "Logs with Zeros? Some Problems and Solutions," *The Quarterly journal of economics*.

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

Criscuolo, C., R. Martin, H. G. Overman, and J. Van Reenen (2019): "Some Causal Effects of an Industrial Policy," *American Economic Review*, 109, 48–85.

Curtis, E. M., D. G. Garrett, E. Ohrn, K. A. Roberts, and J. C. Suarez Serrato (2025): "Capital Investment and Labor Demand: Evidence from 21st Century Stimulus Policy," *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.

Gabaix, X. (2009): "Power Laws in Economics and Finance," *Annual review of economics*, 1, 255–293.

——— (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.

Harasztosi, P. and A. Lindner (2019): "Who Pays for the Minimum Wage?" *American Economic Review*, 109, 2693–2727.

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*.

Hirvonen, J., A. Stenhammar, and J. Tuhkuri (2025): "Winners and Losers of Technology Grants," *Working Paper*.

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

Korkeamäki, O. and R. Uusitalo (2009): "Employment and wage effects of payroll tax cut – evidence from a regional experiment," *International Tax and Public Finance*, 16, 753–772.

Lichter, A., M. Löffler, I. E. Isphording, T.-V. Nguyen, F. Poege, and S. Siegloch (2025): "Profit Taxation, RD Spending, and Innovation," *American Economic Journal: Economic Policy*, 17, 432–63.

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

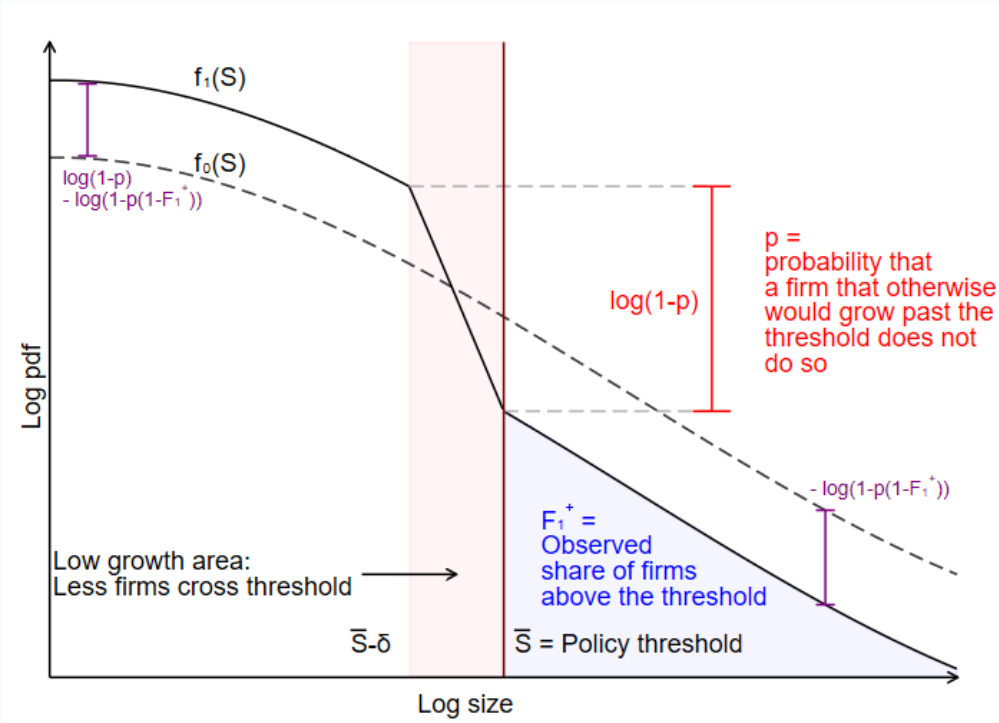
Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023): "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of econometrics*, 235, 2218–2244.

Saez, E., M. Matsaganis, and P. Tsakloglou (2012): "Earnings determination and taxes: evidence from a cohort-based payroll tax reform in Greece," *The Quarterly Journal of Economics*, 127, 493–533.

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.

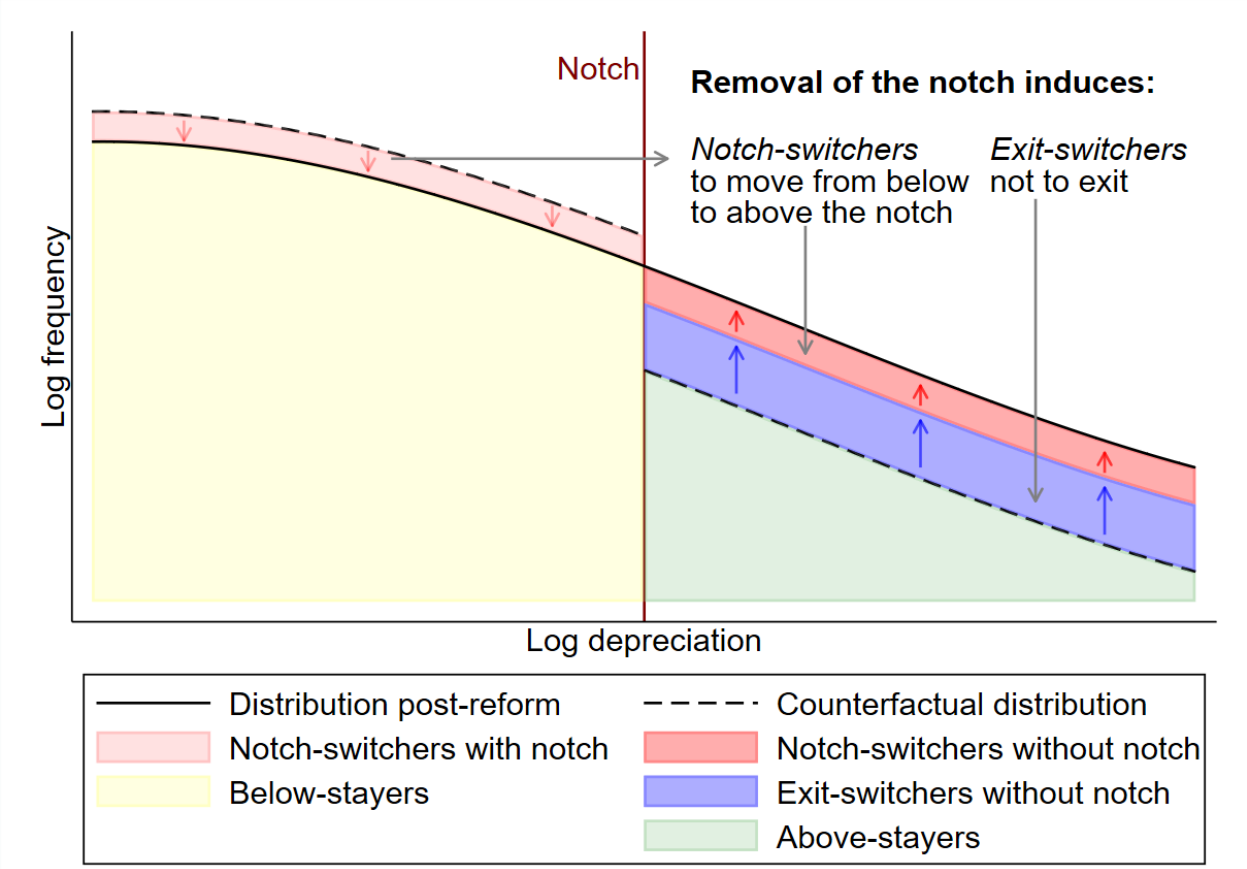
Figures and Tables

Figures



Notes: Figure 1 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 $[\bar{S} - \delta, \bar{S}]$. This would be the case for example if firms do not want to cross the threshold and incur a fixed cost for a small increase in size. p is the probability that a firm that otherwise would have crossed the threshold at $\bar{S}$ does not do so because of the low growth area. 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 by pF_1^+, and the distribution above the threshold down by $p(1 - F_1^+)$. Hence, the shift between the upper and lower branches of the distribution is p. This result holds (approximately) for a large family of dynamic systems of firm growth, including general Itô diffusion and jump processes (see more details in Section 2).

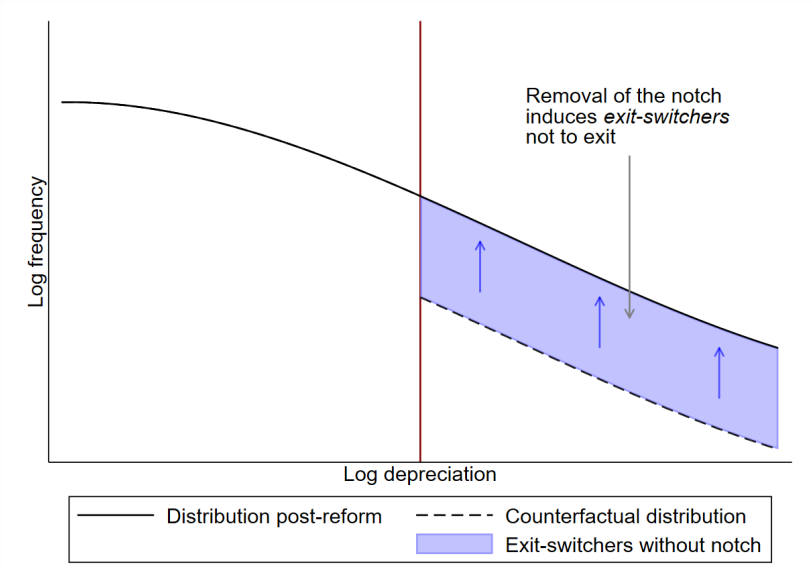
Figure 1: Shifts in the firm distribution due to a growth barrier



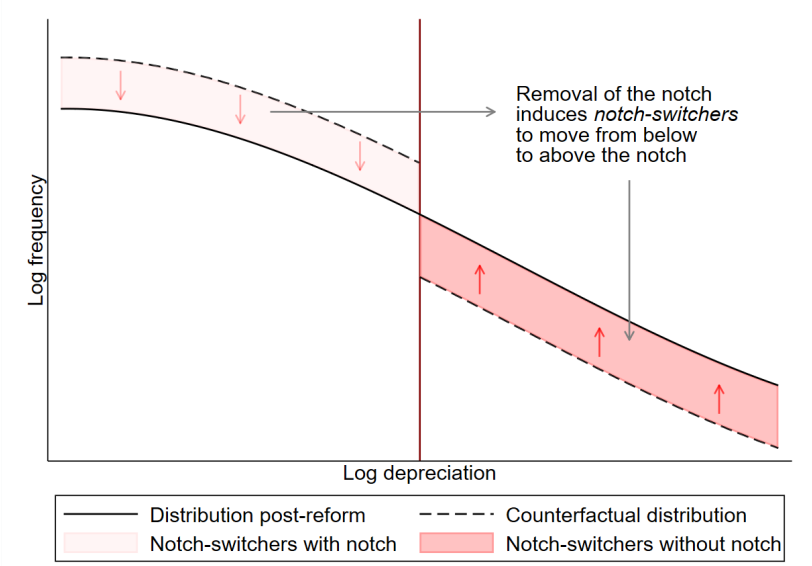
Notes: This figure shows how different groups of firms react to the removal of the notched payroll tax schedule. The notch is located at the red vertical line. The solid black line shows the firm distribution post-reform, i.e. when the notch has been removed. The dashed black line shows the counterfactual distribution had the notch not been removed. The difference between these arises due to notch-switchers and exit-switchers. Removing the notch induces 1) notch switchers to move from below the notch (light red area) to above the notch (red area), and 2) exit-switchers not to exit (blue area). Below-stayers (light yellow area) and above-stayers (light green area) do not react to the removal of the notch, but only above-stayers are defined as treated as they are subject to the treatment. Hence, the treatment group consists of notch-switchers, exit-switchers and above-stayers. Standard diff-in-diff methods assume that there are only below- and above-stayers.

Figure 2: Notch- and exit-switchers

(a) Only exit-switchers

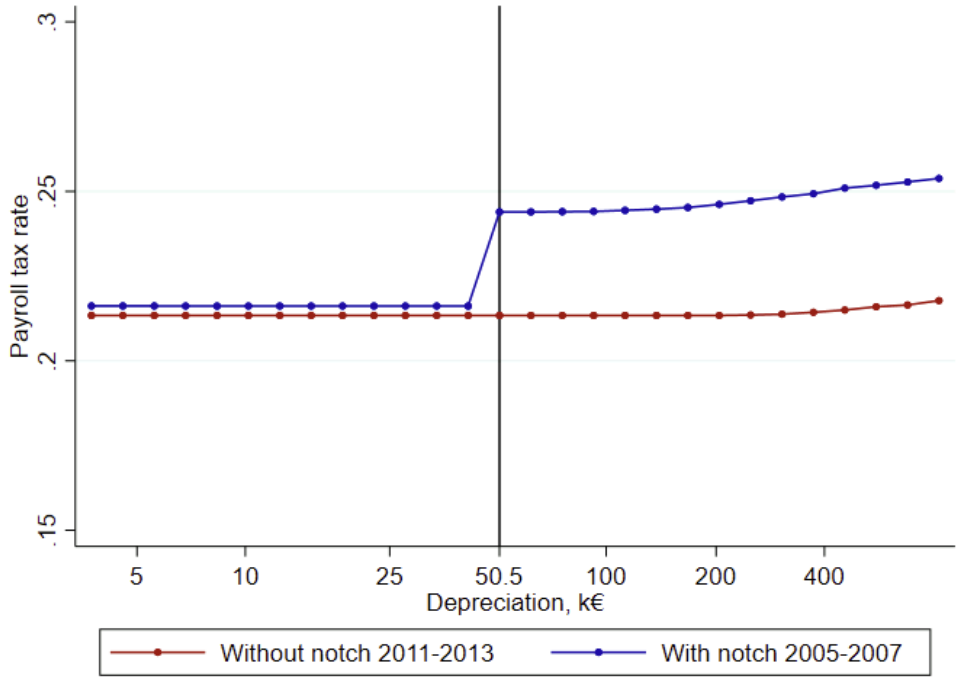


(b) Only notch-switchers



Notes: This figure illustrates how the firm distribution changes with the removal of the notched schedule when there are either only exit-switchers (panel 3a) or only notch-switchers (panel 3b). The solid black line shows the firm distribution post-reform, i.e. when the notch has been removed. The dashed black line shows the counterfactual distribution had the notch not been removed. The notch is located at the red vertical line. When there are only exit-switchers, the two distributions are aligned below the notch. When constructing a counterfactual distribution in this case e.g. from a pre-reform distribution, the counterfactual should be adjusted so that the pre- and post-distributions align below the notch. However, when there are only notch-switchers, the distribution changes on both sides, but the total number of firms does not change. Hence, when constructing a counterfactual distribution in this case, the counterfactual should be adjusted so that the total number of firms is equal to the number of firms post reform.

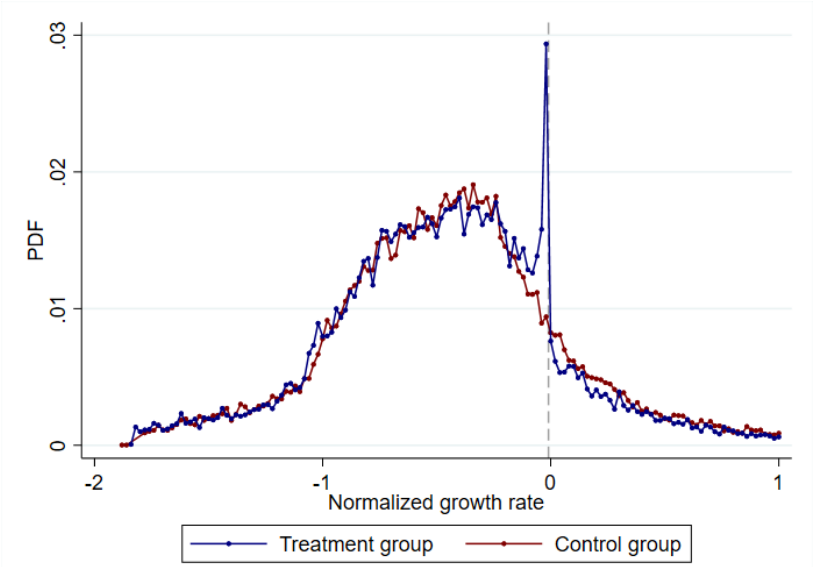
Figure 3: Distributions with and without the notch with only one type of switchers



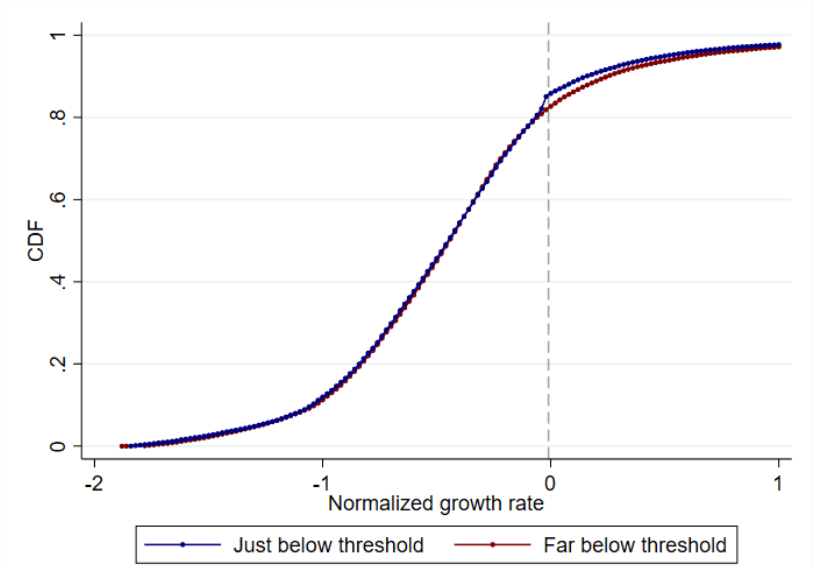
Notes: Figure 4 plots the variation of payroll tax rates in bins with (2005–2007) and without (2011–2013) the notch across the capital depreciation distribution. The width of the bins is calculated such that each bin captures 20 log points of capital depreciation up and down starting from the threshold. 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 4: Payroll Tax Rate by Depreciation Bin: Pre- vs Post-2010

(a) PDF

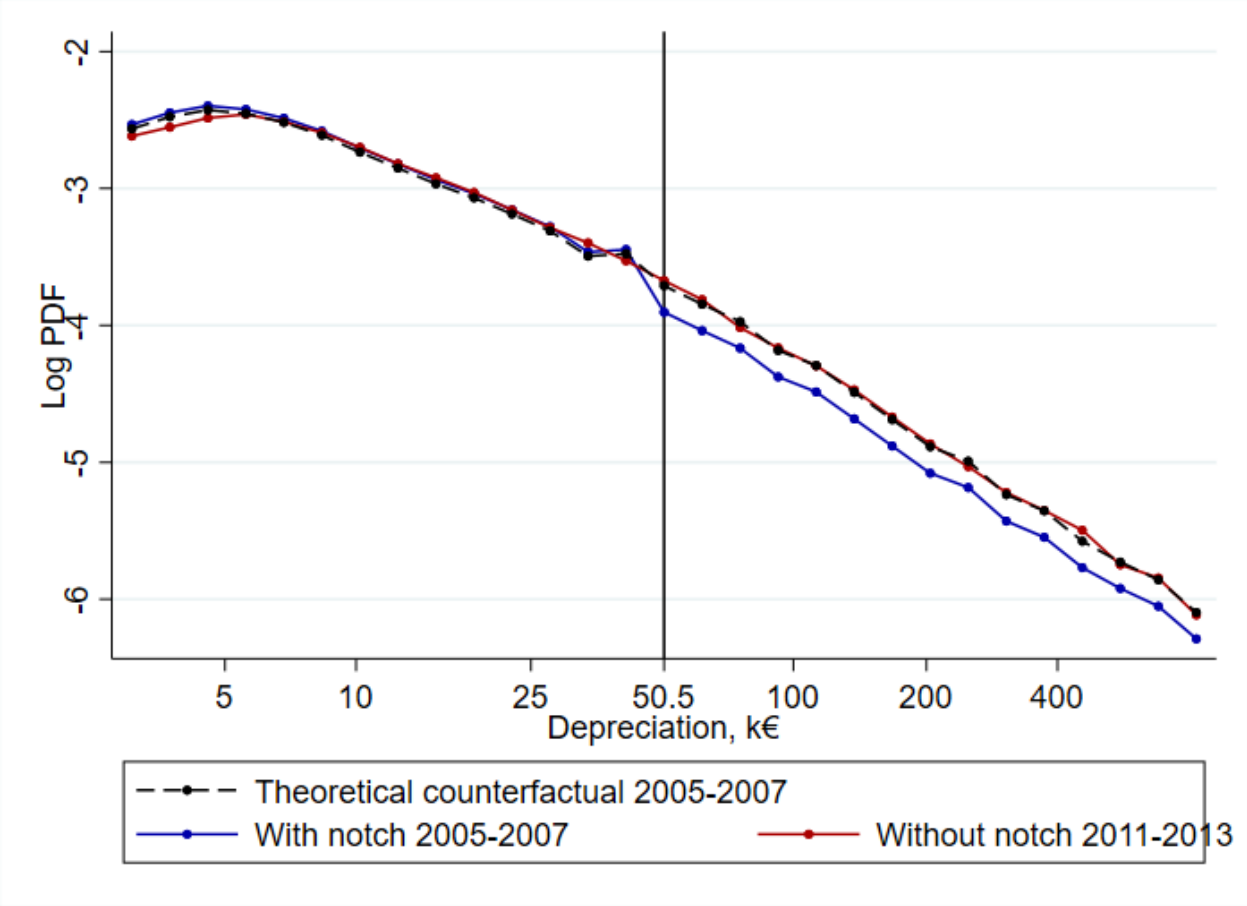


(b) CDF



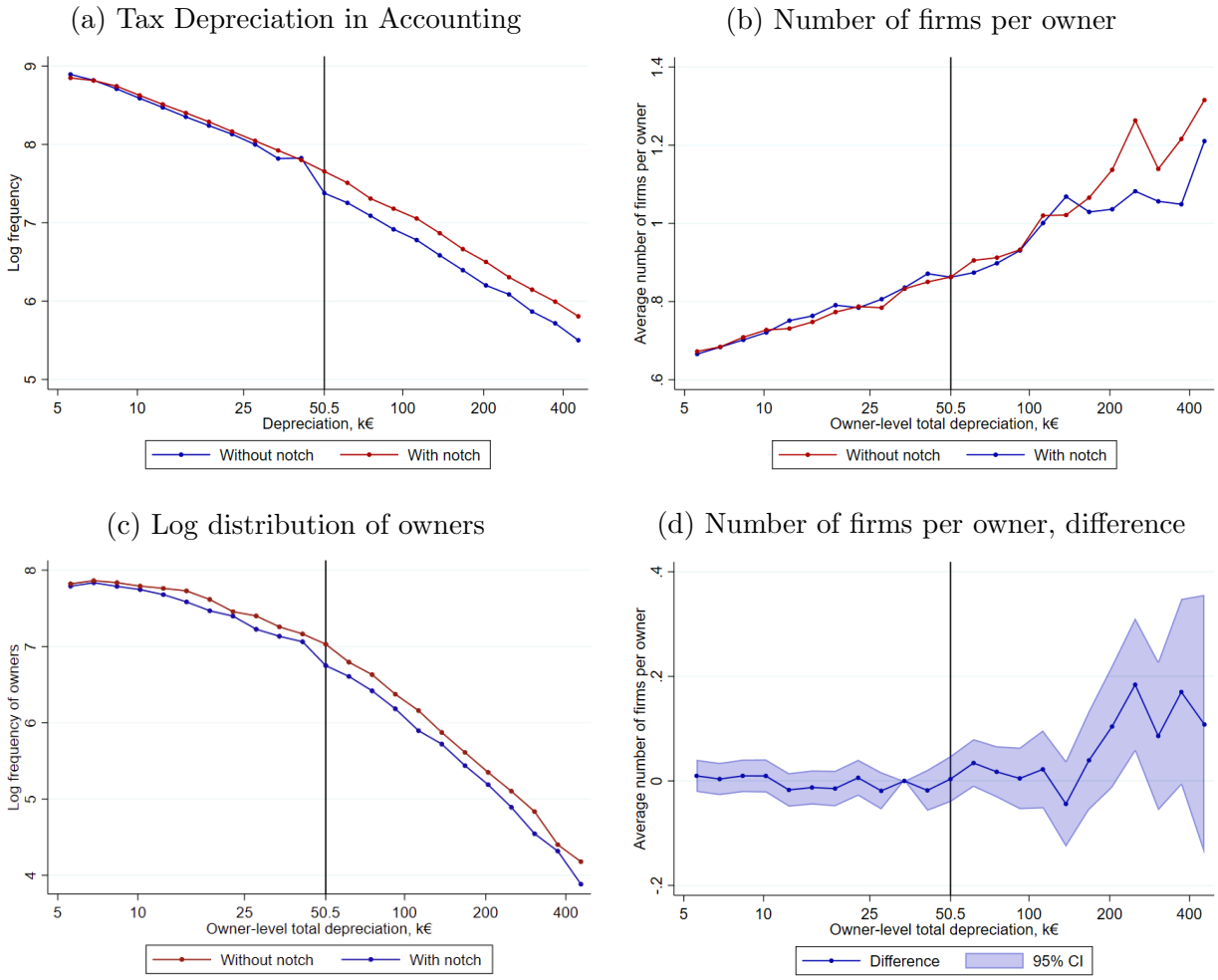
Notes: This figure shows the normalized growth rate distributions of the treated and control groups such that the distributions of the control group have been scaled by a scaling parameter of 0.535 to provide a counterfactual distribution for the treated group. The grey vertical line shows the position of the threshold. Bins have a width of 2%. Panel 5a shows the pdf's and panel 5b shows the cdf's. Treatment group bunching at the threshold is clearly visible in both 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 where the threshold would not exist. The estimate for the probability that a firm that otherwise would have grown past the threshold does not do so is $p = 0.182$ with a bootstrapped standard error of 0.001.

Figure 5: Dynamic bunching distributions



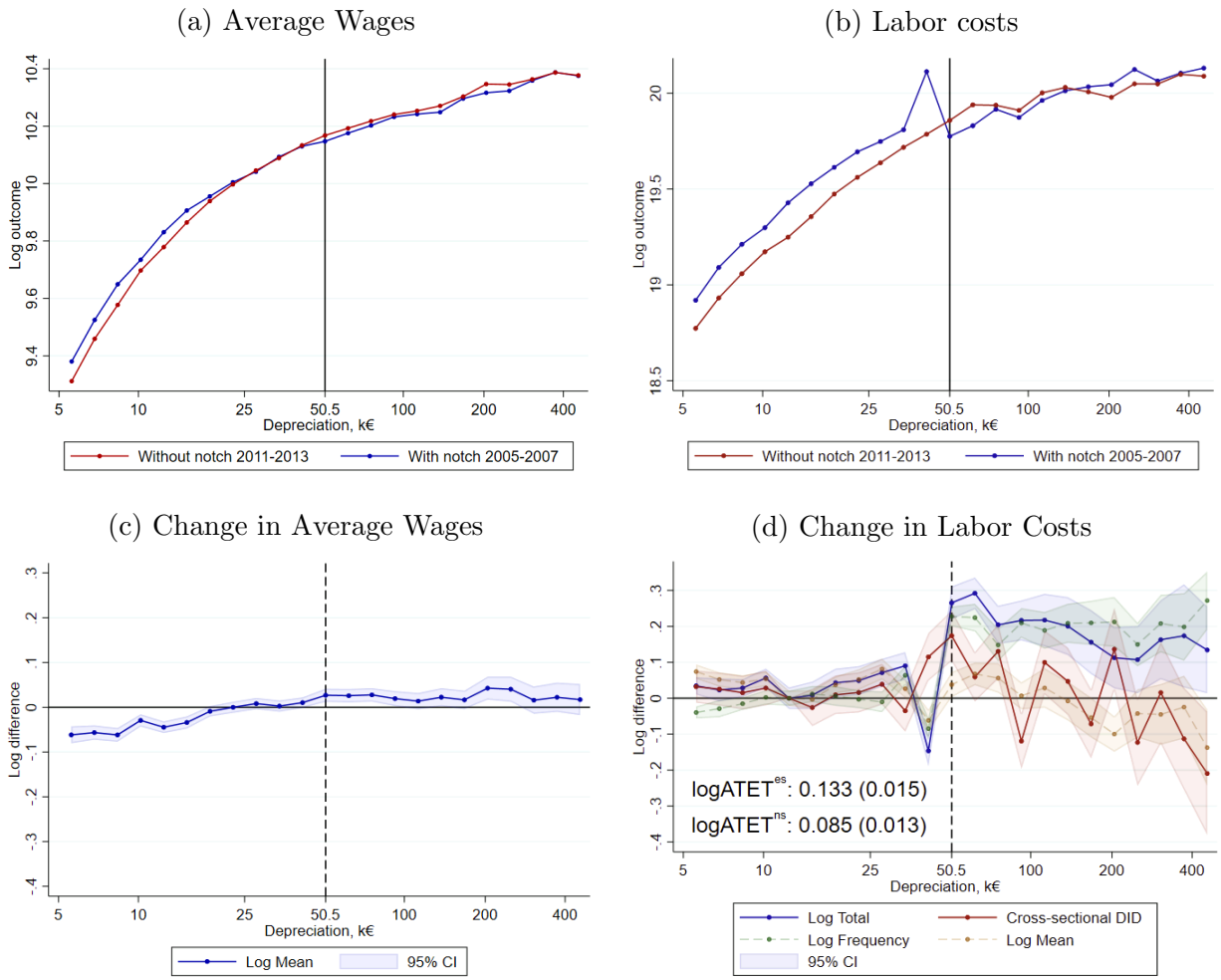
Notes: This figure plots the log pdf's of firms in bins with (2005–2007) and without (2011–2013) the notch across the capital depreciation distribution, and compare these to the theoretical counterfactual calculated based solely on pre-period (2005-2007) data: 1) the firm distribution with the notch (2005–2007), 2) the dynamic bunching estimate of the probability that a firm that otherwise would have grown past the threshold does not do so ($p = 0.182$), and 3) the share of firms above the threshold with the threshold in place (2005–2007). 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 the theoretical counterfactual matches the post-period firm distribution well, providing support for the theoretical result.

Figure 6: Constructed counterfactual and before-after tax notch firm distributions



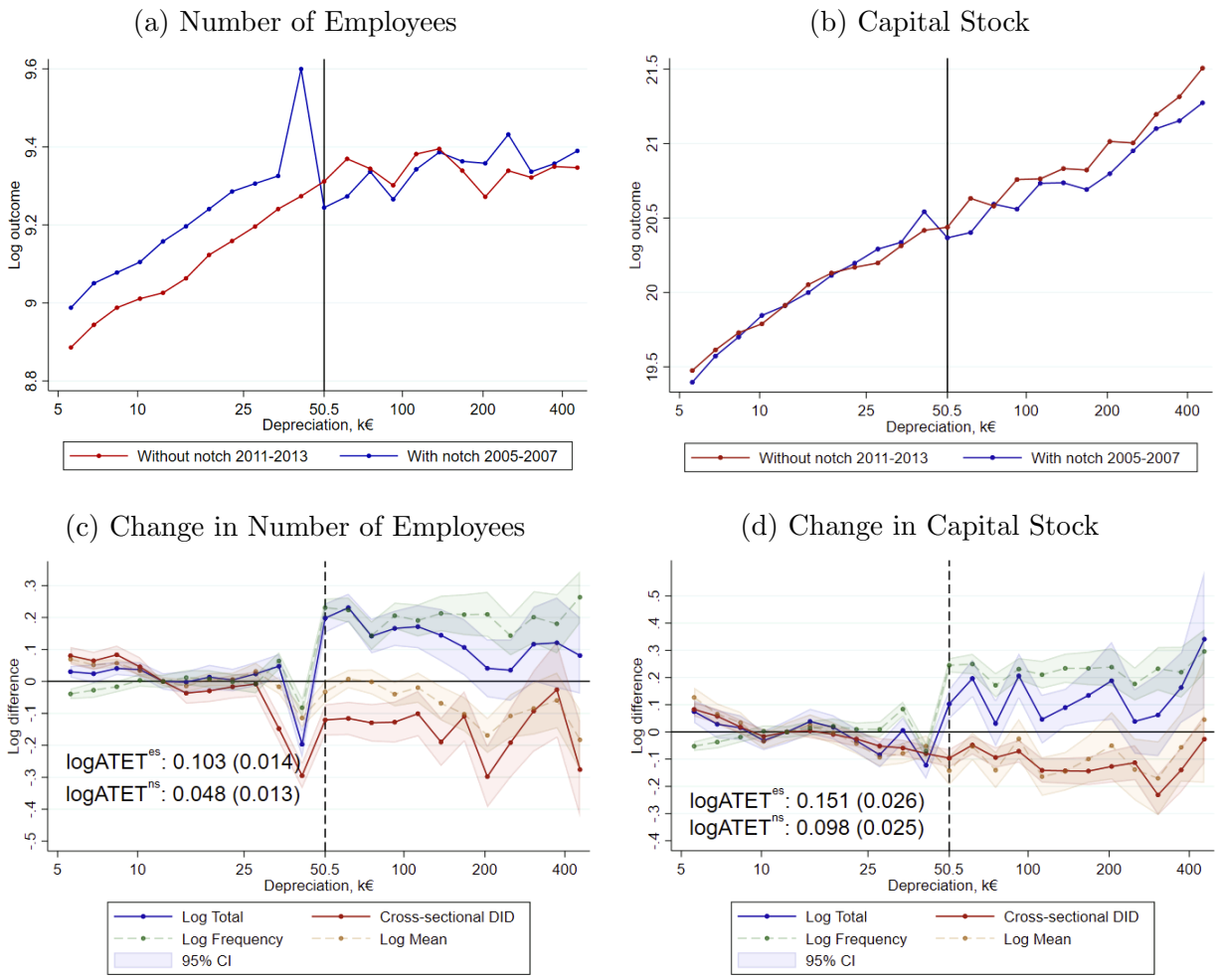
Notes: Figure 7a plots the log frequency of firms using the tax measure of capital depreciation as a running variable with (2005–2007) and without notch (2011–2013). The panel 7b shows the mean number of firm per owner within each bin with (2003–2007) and without notch (2011–2015), and the panel 7d depicts the difference of these with 95% confidence intervals. Panel 7c 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 7: Evidence of Avoidance and Evasion Responses



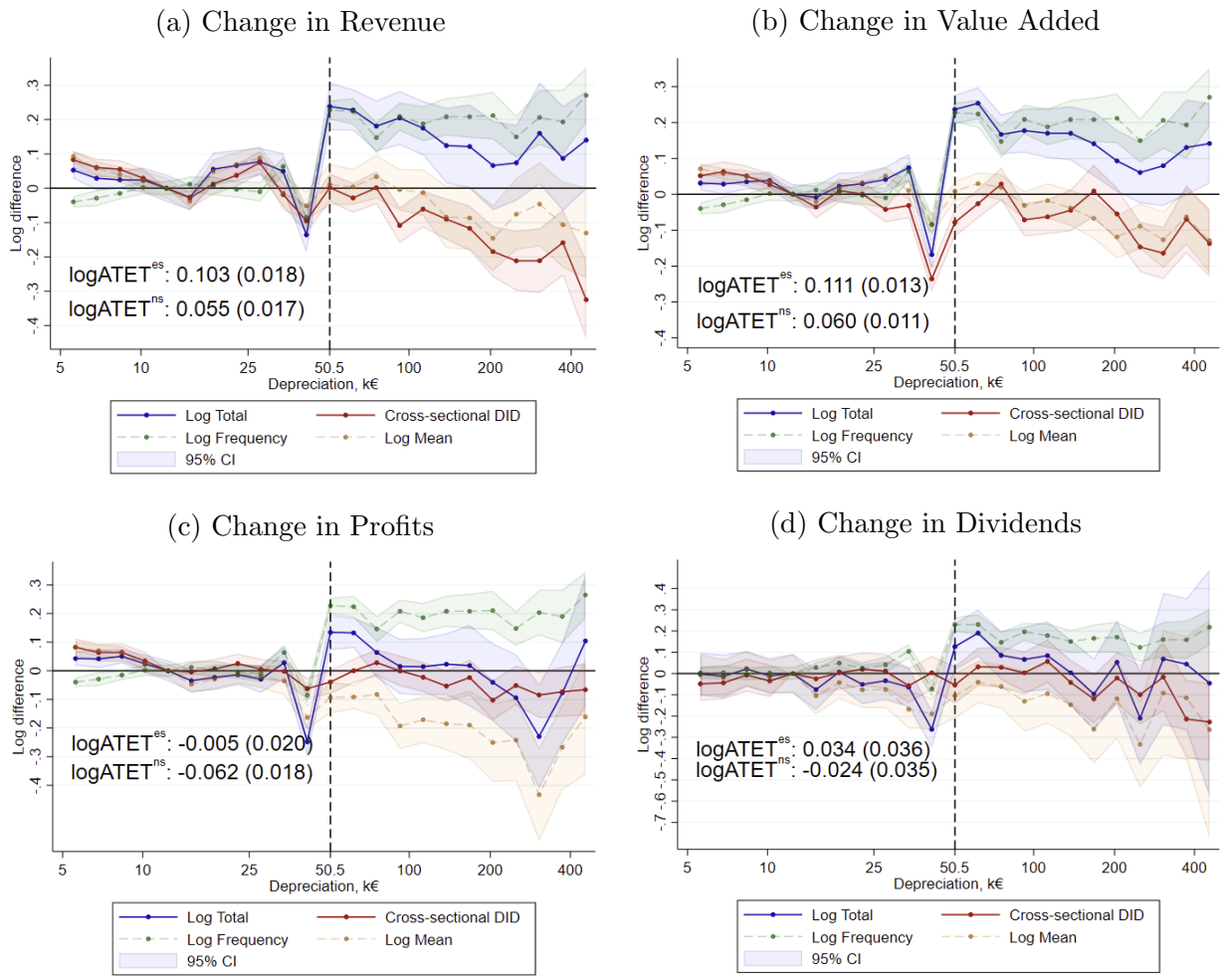
Notes: Panel 8a illustrates the evolution of log mean wage per worker in firms with (2005–2007) and without notch (2011–2013) across the capital depreciation distribution, while panel 8c illustrates difference of these at the bin-level. Panel 8b shows the development for log total labor costs, while panel 8d shows the bin level log $ATET^{es}$ estimates of Equation (26) in blue and the two components of this. Panel 8d also shows standard repeated cross-sections difference-in-difference estimates in red for comparison, as well as the point estimates of log $ATET^{es}$ and log $ATET^{ns}$ from Equations (27) and (28) (the bin right below the threshold and the one to its right are excluded to reduce the effect of bunching on the estimates). 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 them (see more details on the institutional setup in Section 4.1)

Figure 8: Average Wages and Labor Costs by Depreciation Bins: Pre- vs Post-2010



Notes: Panel 9a illustrates the evolution of log total number of workers in each bin with (2005–2007) and without notch (2011–2013) across the capital depreciation distribution, while panel 9c shows the bin level log $ATET^{es}$ estimates of Equation (26) for the same outcome in blue and the two components of this. Panel 9c also shows standard repeated cross-sections difference-in-difference estimates in red for comparison, as well as the point estimates of log $ATET^{es}$ and log $ATET^{ns}$ from Equations (27) and (28) (the bin right below the threshold and the one to its right are excluded to reduce the effect of bunching on the estimates). Panels 9b and 9d show similar estimates for the capital stock of the firm. 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 them (see more details on the institutional setup in Section 4.1)

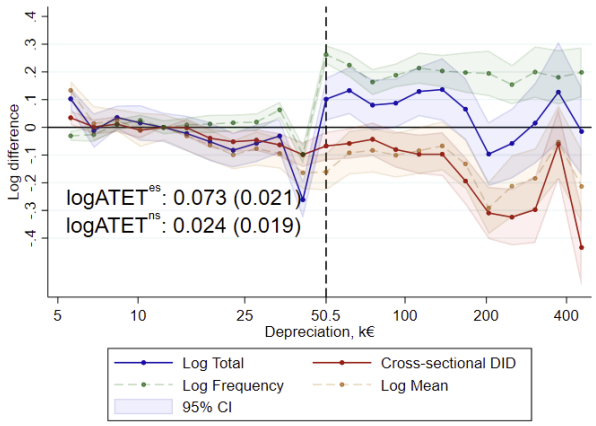
Figure 9: Employment and Capital Stock by Depreciation Bins: Pre- vs Post-2010



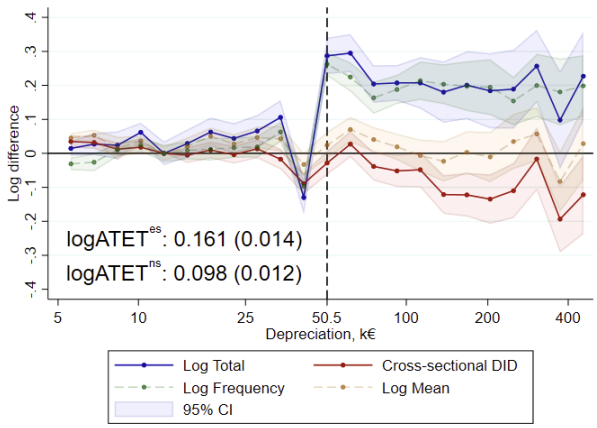
Notes: Panel 10a shows the bin level $\log ATET^{es}$ estimates of Equation (26) for revenue in blue and the two components of this. The panel also shows standard repeated cross-sections difference-in-difference estimates in red for comparison, as well as the point estimates of $\log ATET^{es}$ and $\log ATET^{ns}$ from Equations (27) and (28) (the bin right below the threshold and the one to its right are excluded to reduce the effect of bunching on the estimates). Panels 10b, 10c, and 10d show similar estimates for value added, profits and dividends. 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 10: Revenue, Value Added, Profits and Dividends by Depreciation Bins: Pre- vs Post-2010

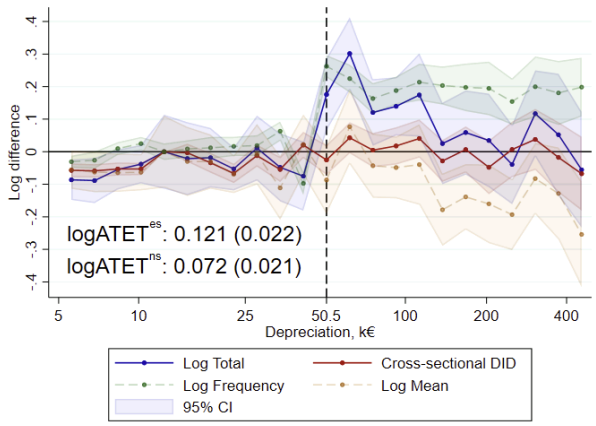
(a) Change in No. of Routine Employees



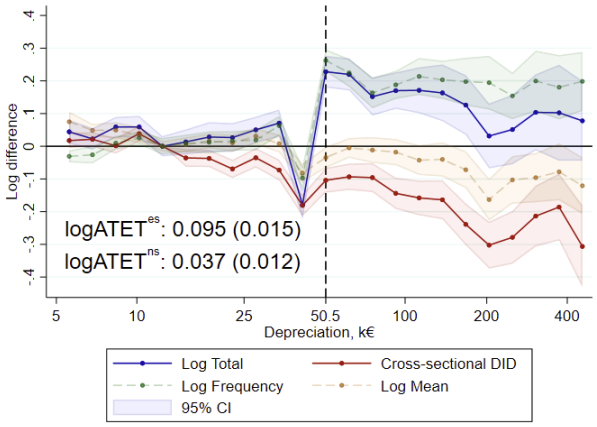
(b) Change in No. of Non-routine Employees



(c) Change in No. of Graduate Employees



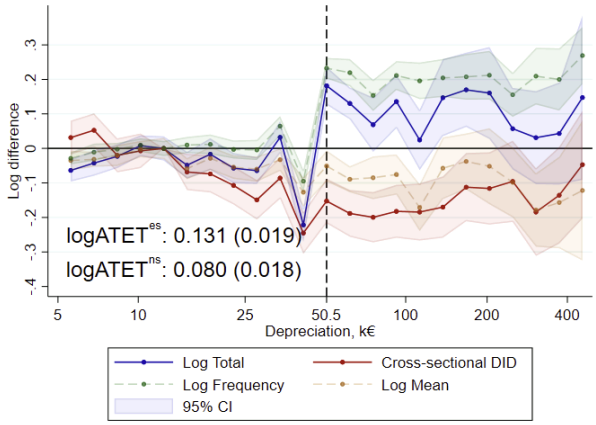
(d) Change in No. of Non-Graduate Employees



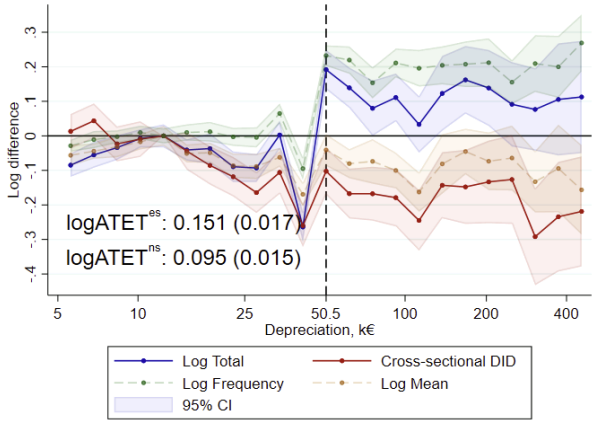
Notes: Panel 11a shows the bin level log $ATET^{es}$ estimates of Equation (26) for routine employees in blue and the two components of this. The panel also shows standard repeated cross-sections difference-in-difference estimates in red for comparison, as well as the point estimates of log $ATET^{es}$ and log $ATET^{ns}$ from Equations (27) and (28) (the bin right below the threshold and the one to its right are excluded to reduce the effect of bunching on the estimates). Panels 11b, 11c, and 11d show similar estimates for non-routine employees, graduate employees, and non-graduate employees. 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 11: Employment by Occupation Group by Depreciation Bins: Pre- vs Post-2010

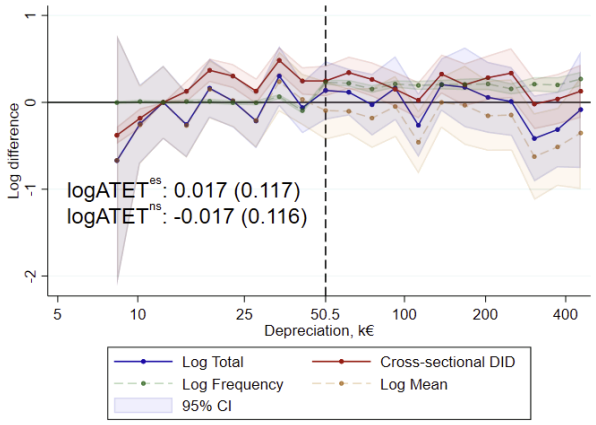
(a) Change in Total Investment



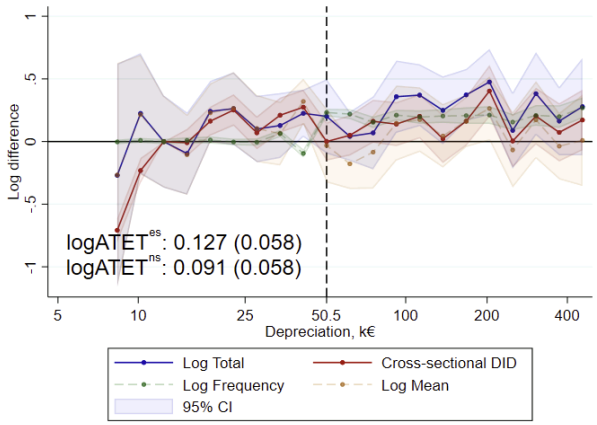
(b) Change in Investment in Fixed Assets



(c) Change in Investment in Buildings

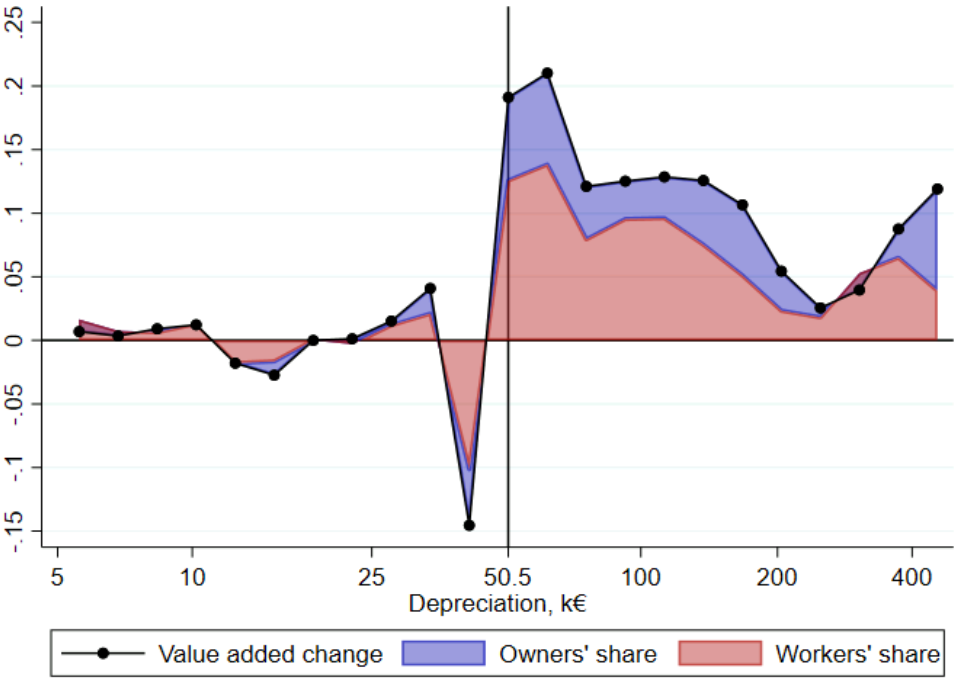


(d) Change in Investment in R&D



Notes: Panel 12a shows the bin level log $ATET^{es}$ estimates of Equation (26) for investment in blue and the two components of this. The panel also shows standard repeated cross-sections difference-in-difference estimates in red for comparison, as well as the point estimates of log $ATET^{es}$ and log $ATET^{ns}$ from Equations (27) and (28) (the bin right below the threshold and the one to its right are excluded to reduce the effect of bunching on the estimates). Panels 12b, 12c, and 12d show similar estimates for investment in fixed assets, buildings and R&D. 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 12: Investment by Type by Depreciation Bins: Pre- vs Post-2010



Notes: Figure 13 illustrates the three terms in Equation (31). The thick black line shows differences in total value added within bins without (2011–2013) versus with (2005–2007) the notch regime divided by pre-treatment (2005–2007) total value added within the same bin. The change in value added is divided into owners' and workers' shares.

Figure 13: Incidence Shares

Tables

	All	Sample $5k < b < 450k$	Last b on the left of the threshold
Capital Depreciation in taxation	62,772	39,979	44,614
Capital Depreciation in accounting	45,922	36,554	46,207
Labor Costs	230,818	153,128	208,647
Number of Employees	7.69	5.36	6.71
Routine	3.30	2.19	2.57
Non-Routine	4.16	3.02	3.96
Graduate	1.24	0.50	0.50
Non-Graduate	6.45	4.86	6.22
Male	4.81	3.74	4.79
Female	2.81	1.57	1.86
Revenue	1,336,311	697,635	875,322
Value added	323,297	216,373	288,929
Capital stock	896,150	337,081	329,762
Total Investment	94,165	64,387	87,824
Investment in fixed assets	49,351	45,291	69,368
Investment in buildings	13,272	9,886	10,130
Investment in R&D	28,356	5,011	3,580
Rental Costs	30,862	14,419	16,927
Services bought outside	97,117	39,414	41,041
Dividends	98,702	23,375	23,248
Number of firms	702,103	180,321	7,811
Number of owners			

Notes: Table 1 presents the mean value of each variable used in our analysis for three different samples: All firms, our restricted sample and last bin on the left of the threshold in our pre-reform period (i.e. years 2005-2007). The restricted sample includes firms that had capital depreciation between 5k€ and 450k€, and 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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Labor Cost	Average Wages	Number of Employees	Capital Stock	Revenue	Value Added	Taxable Profits	Dividends
Overall change with full sample								
$\log ATET^{es}$	0.133		0.103	0.151	0.103	0.111	-0.005	0.034
	(0.015)		(0.014)	(0.026)	(0.018)	(0.013)	(0.020)	(0.036)
$\log ATET^{ns}$	0.085		0.048	0.098	0.055	0.060	-0.062	-0.024
	(0.013)		(0.012)	(0.025)	(0.017)	(0.011)	(0.018)	(0.035)
Log no. firms change with full sample								
τ_n^{DID}	0.206		0.203	0.227	0.205	0.205	0.204	0.154
	(0.010)		(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Log mean change with full sample								
τ^{DID}	-0.073	0.042	-0.100	-0.076	-0.102	-0.094	-0.209	-0.121
	(0.010)	(0.004)	(0.010)	(0.024)	(0.015)	(0.009)	(0.018)	(0.035)
Adjustment term								
Ψ	-0.048		-0.055	-0.053	-0.047	-0.051	-0.057	-0.057
	(0.002)		(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Observations	342,712	339,032	339,032	313,886	342,746	342,746	342,707	188,072
Overall change with depr. 12.5-125k								
$\log ATET^{es}$	0.184		0.161	0.133	0.152	0.160	0.065	0.118
	(0.018)		(0.017)	(0.024)	(0.021)	(0.016)	(0.024)	(0.034)
$\log ATET^{ns}$	0.111		0.081	0.057	0.080	0.083	-0.019	0.037
	(0.014)		(0.014)	(0.021)	(0.018)	(0.013)	(0.022)	(0.032)
Log no. firms change with depr. 12.5-125k								
τ_n^{DID}	0.184		0.180	0.197	0.183	0.183	0.182	0.153
	(0.012)		(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.015)
Log mean change with depr. 12.5-125k								
τ^{DID}	-0.000	0.029	-0.019	-0.064	-0.031	-0.023	-0.118	-0.035
	(0.012)	(0.005)	(0.012)	(0.020)	(0.017)	(0.011)	(0.023)	(0.033)
Adjustment term with depr. 12.5-125k								
Ψ	-0.073	0.043	-0.080	-0.076	-0.073	-0.076	-0.084	-0.081
	(0.005)	(0.001)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)	(0.008)
Observations	163,227	161,438	161,438	153,873	163,240	163,240	163,227	102,749

Notes: Table 2 provides estimates of $\log ATET$ for different outcomes using specifications (27) and (28). The table also provides estimates of the different components of these $\log ATET$ estimates. The estimates compare pre (2005-2007) and post (2011-2013) reform outcomes below and above the threshold. We exclude the closest bin below and above the threshold to reduce the effect of bunching on the estimates. The estimates are provided for two different samples: the full sample is the same sample as shown in the bin-level graphs (i.e. firms with depreciation between 4.5k€ and 450k€), and another sample including firms with depreciation between 12.5k€ and 125k€. 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).

Table 2: Reduced-Form Results

	(1)	(2)	(3)	(4)	(5)	(6)
	Routine	Non-Routine	Graduate	Non-Graduate	Male	Female
Full sample						
Change in share	-2.28	2.28	1.19	-1.19	-0.40	0.40
	(0.48)	(0.48)	(0.19)	(0.19)	(0.42)	(0.42)
Baseline in treatment group	48.49	51.51	10.91	89.09	70.42	29.58
	(0.30)	(0.30)	(0.11)	(0.11)	(0.24)	(0.24)
Observations	269,958	269,958	269,958	269,958	269,958	269,958
Sample with depr. 12.5-125k						
Change in share	-0.70	0.70	1.06	-1.06	-0.69	0.69
	(0.63)	(0.63)	(0.25)	(0.25)	(0.54)	(0.54)
Baseline in treatment group	42.39	57.61	8.71	91.29	71.63	28.37
	(0.34)	(0.34)	(0.14)	(0.14)	(0.34)	(0.34)
Observations	132,363	132,363	132,363	132,363	132,363	132,363

Notes: Table 3 provides estimates of percentage point changes by employment category (routine vs non-routine, graduate vs. non-graduate, and male vs female) due to the reform. The estimates compare pre (2005-2007) and post (2011-2013) reform outcomes below and above the threshold. We exclude the closest bin below and above the threshold to reduce the effect of bunching on the estimates. The estimates are provided for two different samples: the full sample is the same sample as shown in the bin-level graphs (i.e. firms with depreciation between 4.5k€ and 450k€), and another sample including firms with depreciation between 12.5k€ and 125k€. 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). All standard errors have been bootstrapped at the firm-level.

Table 3: Reduced-Form Results, Compositional Changes By Employee Categories

	(1)	(2)	(3)
	Investment in Fixed Assets	Investment in Buildings	Investment in R&D
Full sample			
Change in share	0.65	-3.38	2.72
	(0.92)	(0.94)	(0.48)
Baseline in treatment group	73.72	15.77	10.51
	(0.67)	(0.70)	(0.34)
Observations	323,466	323,466	323,466
Sample with depr. 12.5-125k			
Change in share	0.19	-1.60	1.42
	(0.95)	(0.78)	(0.56)
Baseline in treatment group	83.48	9.29	7.23
	(0.65)	(0.56)	(0.34)
Observations	154,932	154,932	154,932

Notes: Table 4 provides estimates of percentage point changes by investment category (fixed assets, buildings, R&D) due to the reform. The estimates compare pre (2005-2007) and post (2011-2013) reform outcomes below and above the threshold. We exclude the closest bin below and above the threshold to reduce the effect of bunching on the estimates. The estimates are provided for two different samples: the full sample is the same sample as shown in the bin-level graphs (i.e. firms with depreciation between 4.5k€ and 450k€), and another sample including firms with depreciation between 12.5k€ and 125k€. 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). All standard errors have been bootstrapped at the firm-level.

Table 4: Reduced-Form Results, Compositional Changes By Investment Category

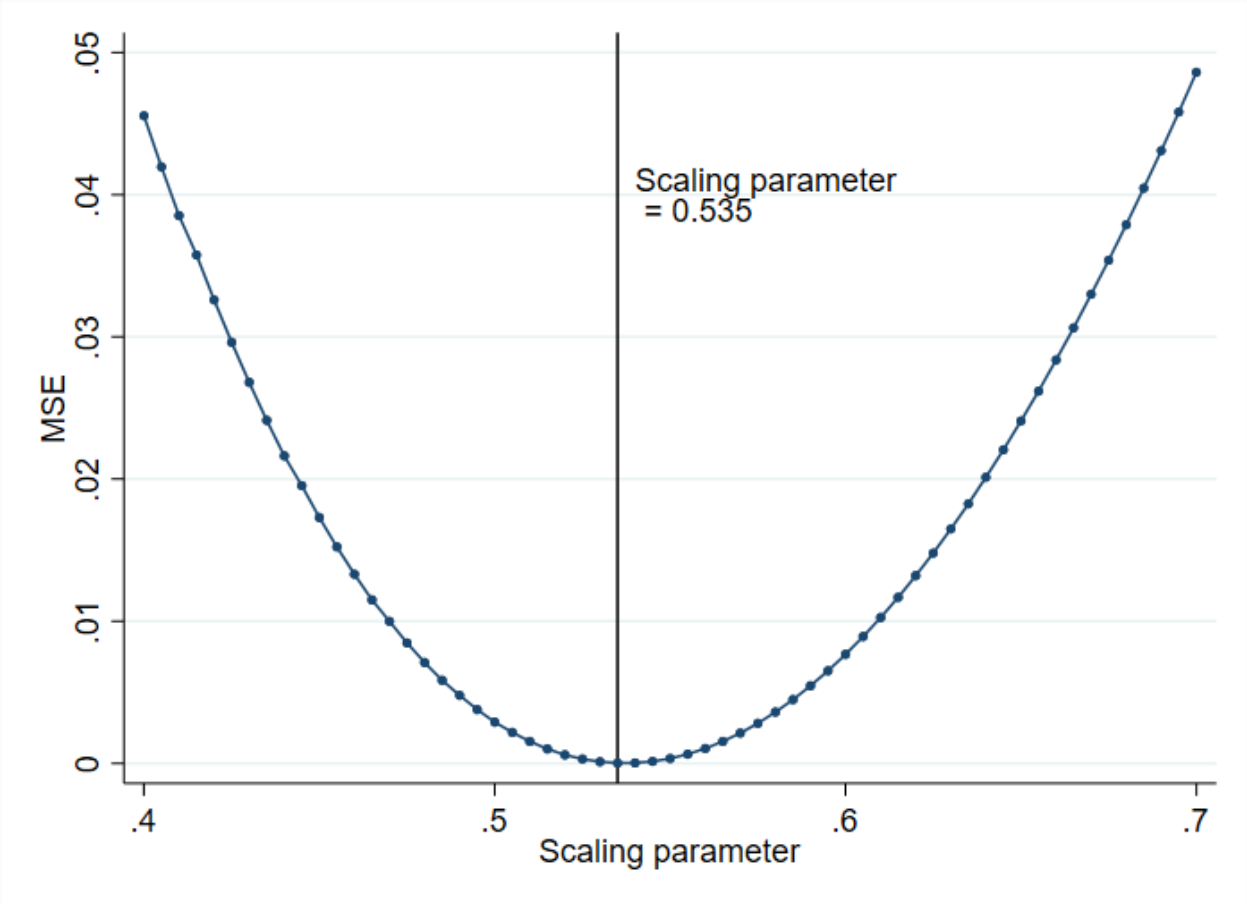
	(1)	(2)	(3)
	$\frac{\Delta VA}{VA_{pre}}$	$\frac{\Delta(\Pi+rK)}{VA_{pre}}$	$\frac{\Delta(wL)}{VA_{pre}}$
	Value Added	Owners	Workers
Change (percent)	9.584	3.525	6.059
	(1.269)	(0.544)	(0.943)
Incidence share (percent)		36.778	63.222
		(3.919)	(3.919)
Observations	342,746	342,746	342,746

Notes: Table 5 provides the estimates separately for the three incidence measures presented in Equation (31) and provides also estimates for the share of incidence between capitalists and workers.

Table 5: Incidence Results

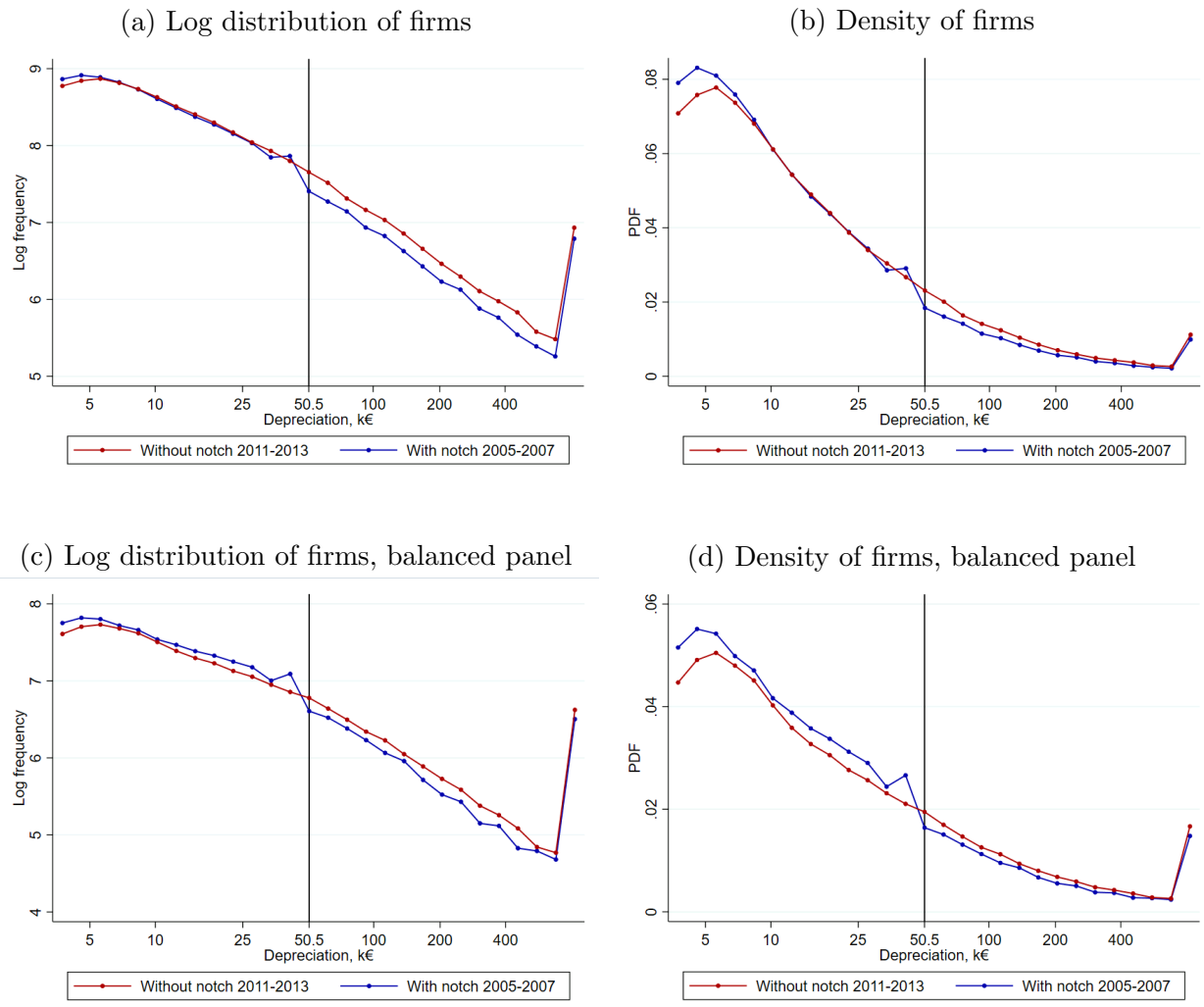
A Appendix

Figures



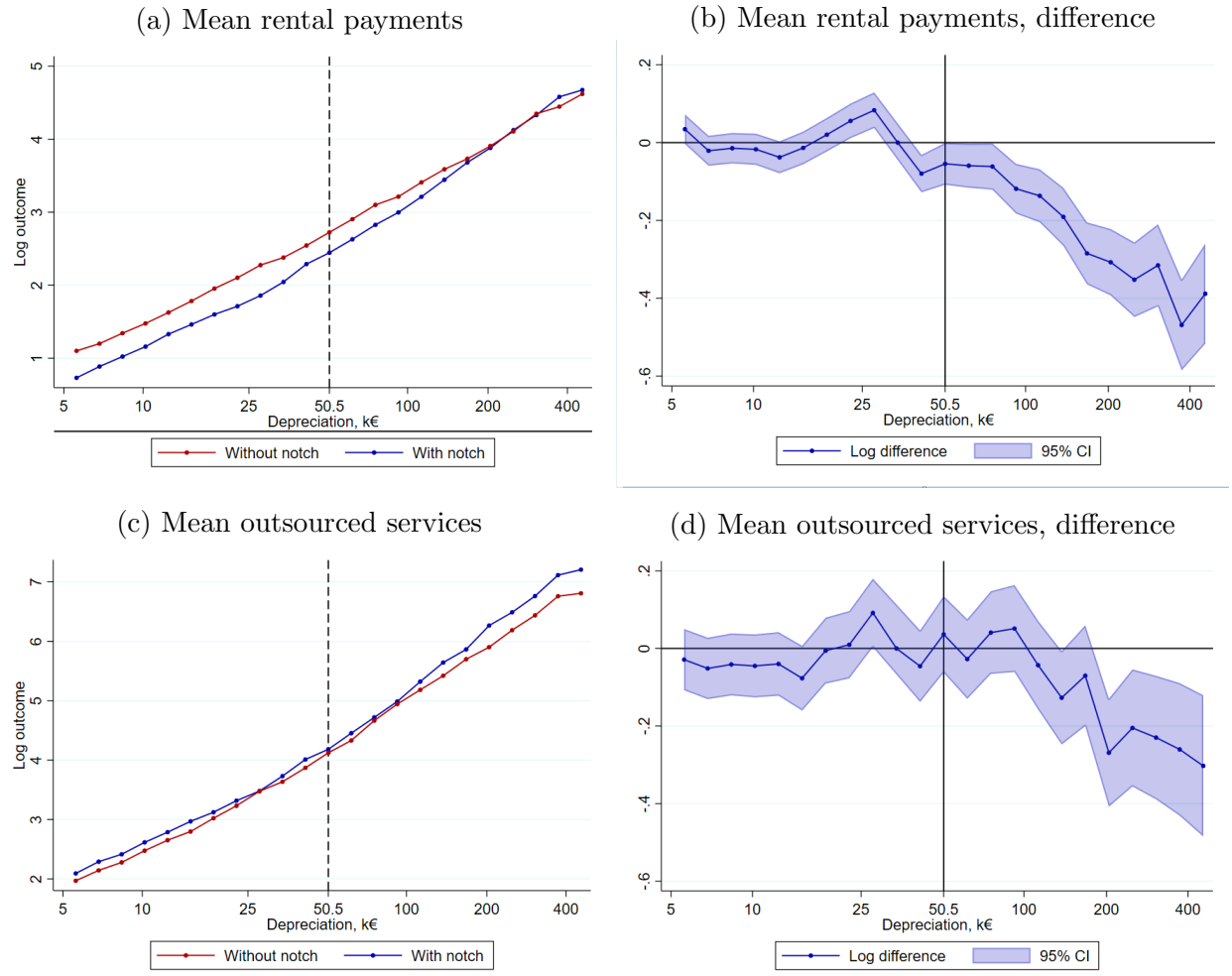
Notes: This figure shows the mean squared errors (MSE) of matching the cumulative growth rate distributions of the control group (depreciation between 15,210 € to 27,714 €) to the treated group (depreciation between 22,7715 € to 50,499 €) in our dynamic bunching setup. The distributions are matched for growth rate bins below -4% below the threshold (true threshold for treated group, placebo threshold for control group). The MSE's 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



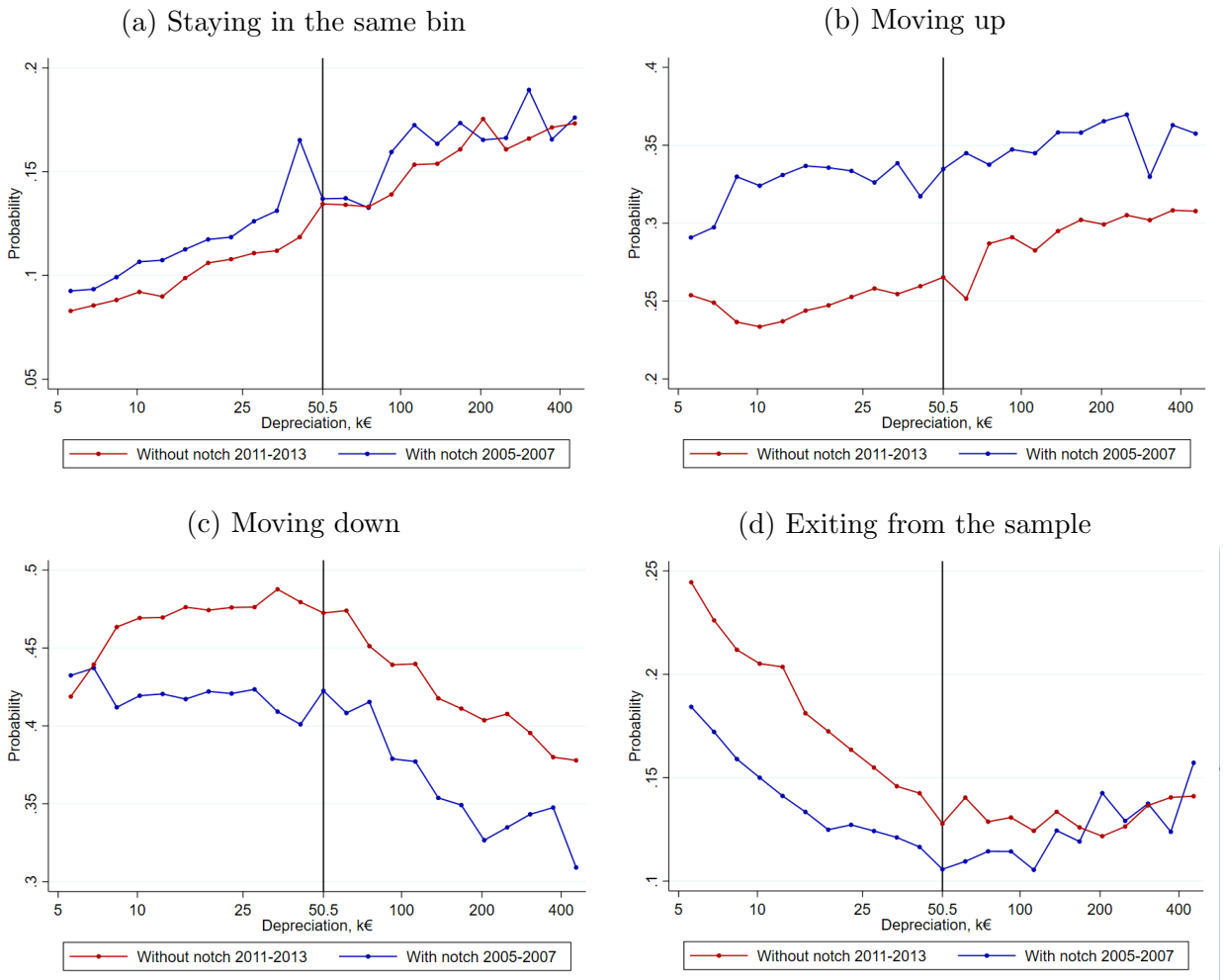
Notes: Figure A2a plots the log frequency of firms in bins with (2005–2007) and without (2011–2013) the notch across the capital depreciation distribution. The width of the bins is calculated such that each bin captures 20 log points of capital depreciation up and down starting from the threshold. Only 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



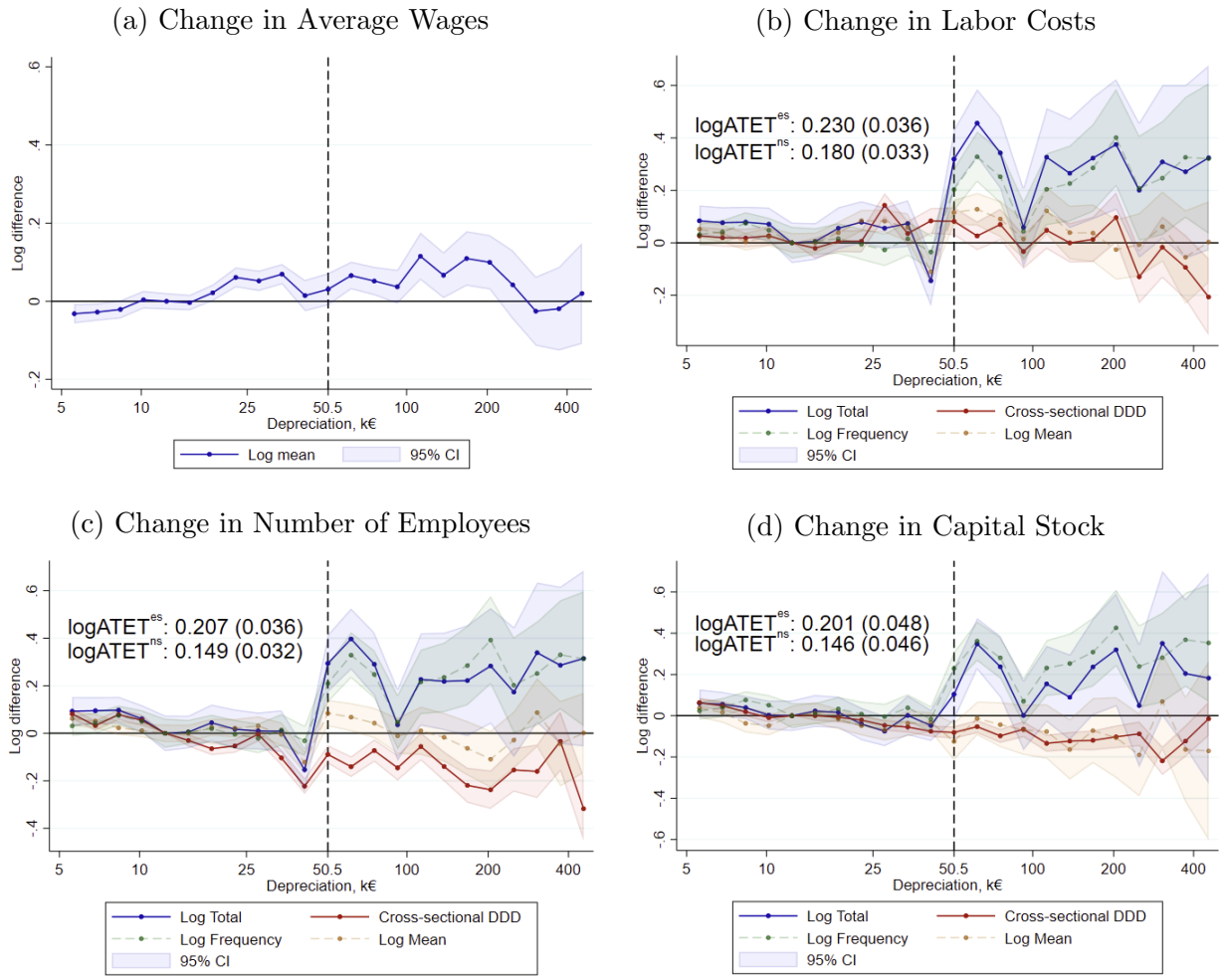
Notes: The panel A3a shows the mean of rental payments within each bin with (2003–2007) and without notch (2011–2015), and the panel A3b depicts the difference of these two regimes with 95% confidence intervals. The panel A3c shows the mean of outsourcing within each bin with (2003–2007) and without notch (2011–2015), and the panel A3b 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 A3: Firms using outsourcing or rental capital to avoid exceeding the threshold



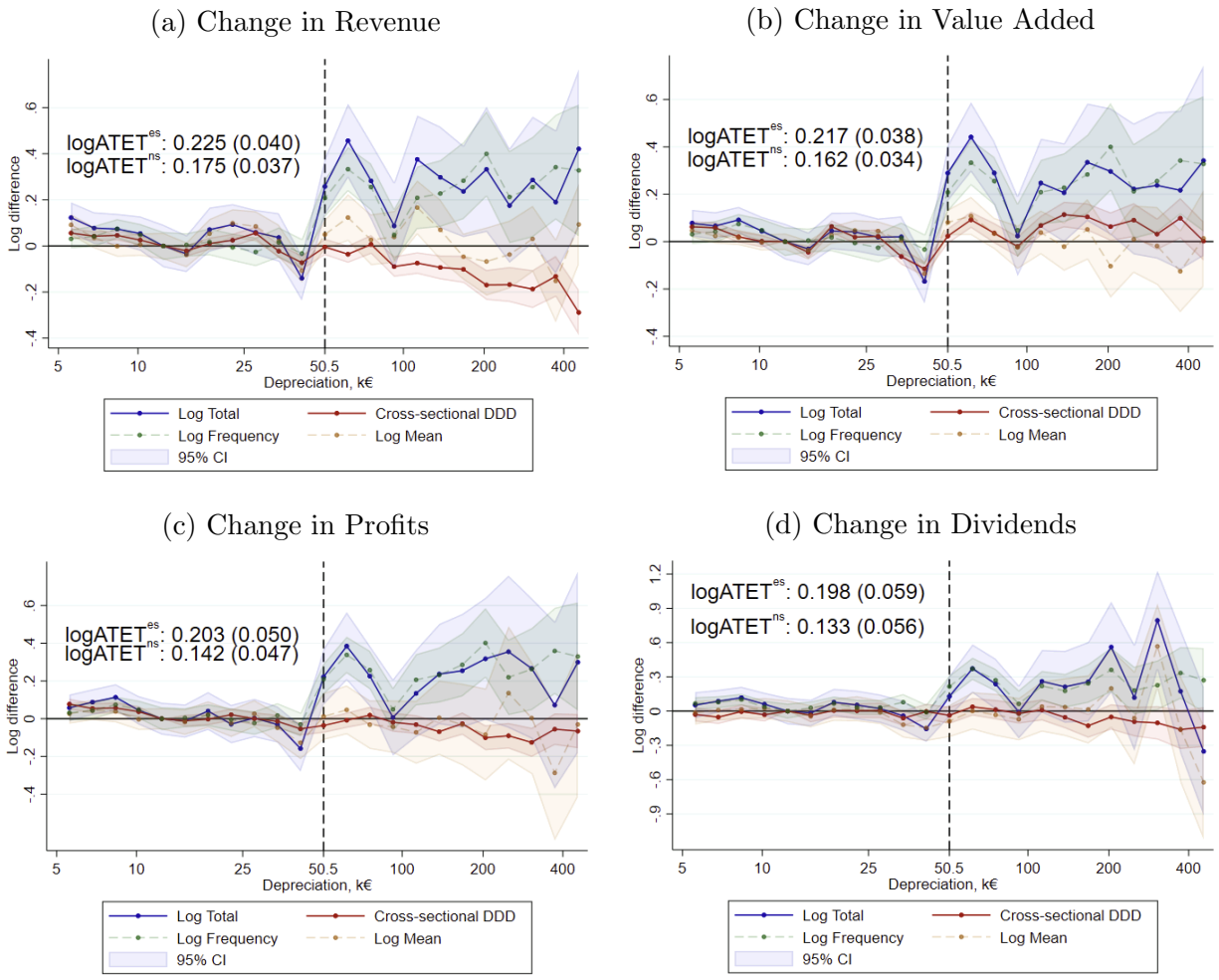
Notes: Figure A4 plots the average likelihood of firms staying in the same bin (Figure A4a), moving some higher bin (Figure A4b), moving to some bin lower than where they were (Figure A4c), or exiting the sample (Figure A4d) between years 2002–2004 and 2005–2007 (with notch), and years 2011–2013 and 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 A4: Firms movement in the distribution.



Notes: Panel A5a illustrates the evolution of log mean wage per worker across the capital depreciation comparing periods with (2005–2007) and without notch (2011–2013) in firms that had depreciation more than 10% of their wage sum versus in those that had less in a triple difference setup distribution. Panel A5b shows the bin level triple difference log $ATET^{es}$ estimates of Equation (26) in blue and the two components of this. The panel also shows standard repeated cross-sections triple difference estimates in red for comparison, as well as the point estimates of log $ATET^{es}$ and log $ATET^{ns}$ from Equations (**??**) and (**??**) (the bin right below the threshold and the one to its right are excluded to reduce the effect of bunching on the estimates). Panels A5c and A5d show the same estimates for the number of employees and the capital stock. 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.

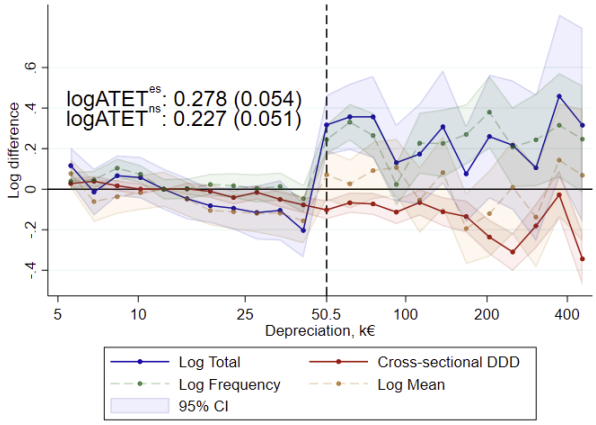
Figure A5: Average Wages, Labor Costs, Employment and Capital Stock by Depreciation Bins: Pre- vs Post-2010, Triple Differences



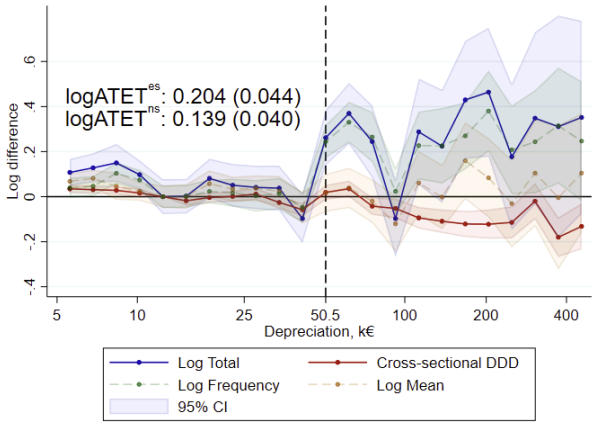
Notes: Panel **??** shows the bin level triple difference log $ATET^{es}$ estimates of Equation (26) for revenue in blue and the two components of this. The additional control group are firms with depreciation less than 10% of their wage sum (see Section 4.1 for more information on the institutional setting). The panel also shows standard repeated cross-sections triple difference estimates in red for comparison, as well as the point estimates of log $ATET^{es}$ and log $ATET^{ns}$ from Equations (**??**) and (**??**) (the bin right below the threshold and the one to its right are excluded to reduce the effect of bunching on the estimates). Panels **??**, **??**, and **??** show similar estimates for value added, profits and dividends. 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 A6: Additional Outcomes by Depreciation Bins: Pre- vs Post-2010, Triple Difference

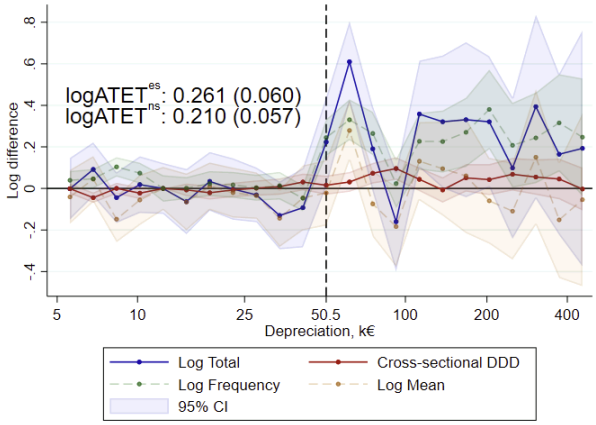
(a) Change in No. of Routine Employees



(b) Change in No. of Non-routine Employees



(c) Change in No. of Graduate Employees



(d) Change in No. of Non-Graduate Employees

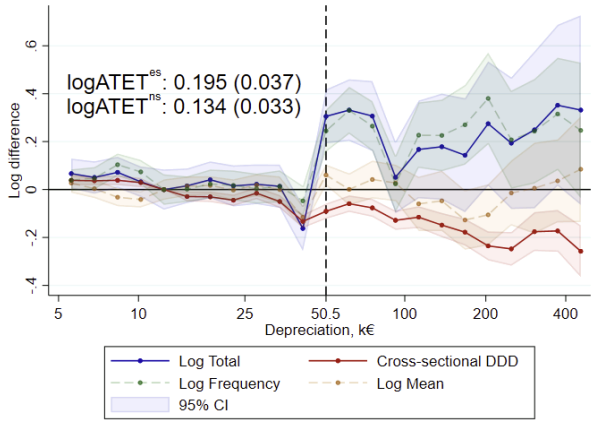
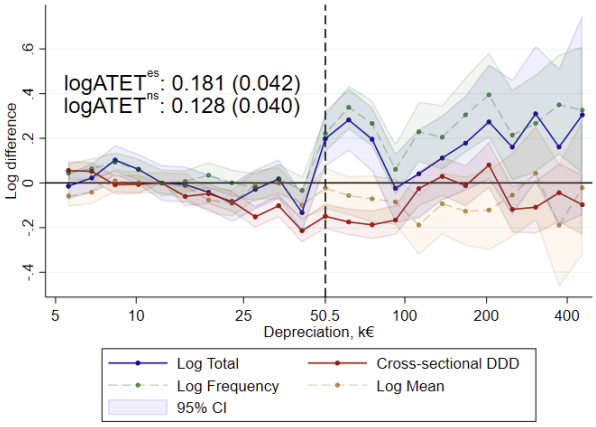


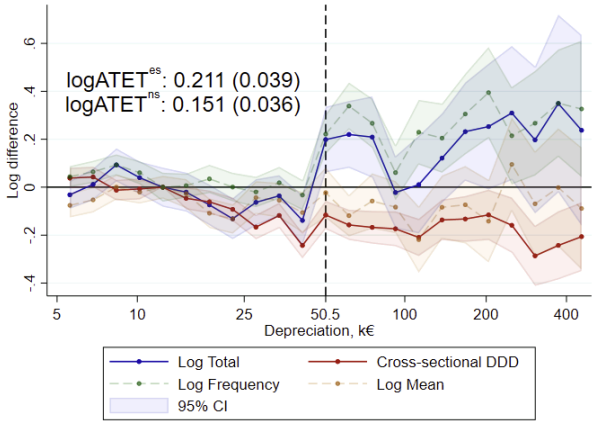
Figure 10 illustrates the ... in Finland. The panels (a) and (c) depicts the ..., while the panels (b) and (d) focus on the ...

Figure A8: Employment by Occupation Group by Depreciation Bins: Pre- vs Post-2010, Triple Differences

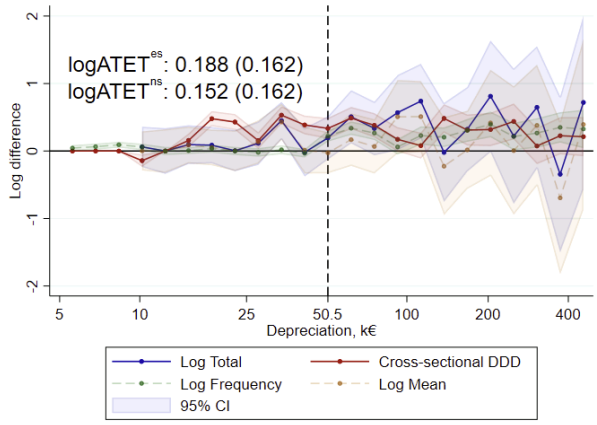
(a) Change in Total Investment



(b) Change in Investment in Fixed Assets



(c) Change in Investment in Buildings



(d) Change in Investment in R&D

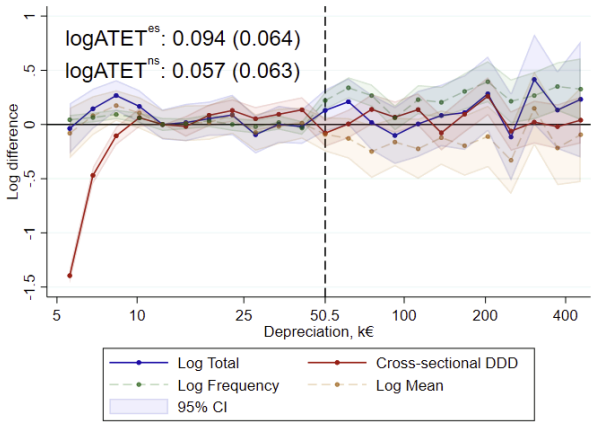


Figure 10 illustrates the ... in Finland. The panels (a) and (c) depicts the ..., while the panels (b) and (d) focus on the ...

Figure A10: Investment by Type by Depreciation Bins: Pre- vs Post-2010, Triple Differences

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 certain wage sums threshold, e.g. 2,059,500 euros in year 2017, and Category II is for wages above the threshold. The threshold varies slightly over years.
*** Represents the average values of these insurances.

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

	(1) Labor Cost	(2) Average Wages	(3) Number of Employees	(4) Capital Stock	(5) Revenue	(6) Value Added	(7) Taxable Profits	(8) Dividends
Overall change with full sample								
$\log ATET^{es}$	0.230		0.207	0.201	0.225	0.217	0.203	0.198
	(0.036)		(0.036)	(0.048)	(0.040)	(0.038)	(0.050)	(0.059)
$\log ATET^{ns}$	0.180		0.149	0.146	0.175	0.162	0.142	0.133
	(0.033)		(0.032)	(0.046)	(0.037)	(0.034)	(0.047)	(0.057)
Log no. firms change with full sample								
$\Delta\tau_n^{DID}$	0.216		0.216	0.240	0.220	0.220	0.222	0.185
	(0.023)		(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Log mean change with full sample								
$\Delta\tau^{DID}$	0.014	0.032	-0.009	-0.039	0.005	-0.003	-0.019	0.013
	(0.025)	(0.013)	(0.022)	(0.044)	(0.028)	(0.025)	(0.040)	(0.054)
Adjustment term								
$\Delta\Psi$	-0.051		-0.058	-0.055	-0.050	-0.054	-0.061	-0.017
	(0.005)		(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
Observations	435,041	431,269	431,269	405,657	435,129	435,129	435,035	269,360

Notes: Table A3 provides estimates of $\log ATET$ for different outcomes using specifications (**??**) and (**??**). The table also provides estimates of the different components of these $\log ATET$ estimates. The estimates compare pre (2005-2007) and post (2011-2013) reform outcomes below and above the threshold in the treated group (firms with depreciation more than 10% of their labor costs) to the same differences in the control group (firms with depreciation less than 10% of their labor costs). We exclude the closest bin below and above the threshold to reduce the effect of bunching on the estimates. The estimates are provided for two different samples: the full sample is the same sample as shown in the bin-level graphs (i.e. firms with depreciation between 4.5k€ and 450k€), and another sample including firms with depreciation between 12.5k€ and 125k€.

Table A3: Reduced-Form Results with an Additional Control Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	No. of employees with education and occupation info	Routine	Non-Routine	Graduate	Non-Graduate	Male	Female
Overall change with full sample							
$\log ATET^{es}$	0.203	0.278	0.204	0.261	0.195	0.215	0.253
	(0.036)	(0.054)	(0.044)	(0.060)	(0.037)	(0.037)	(0.049)
$\log ATET^{ns}$	0.144	0.227	0.139	0.210	0.135	0.157	0.194
	(0.032)	(0.051)	(0.040)	(0.057)	(0.033)	(0.033)	(0.045)
Log no. firms change with full sample							
$\Delta\tau_n^{DID}$	0.200	0.200	0.200	0.200	0.200	0.200	0.200
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Log mean change with full sample							
$\Delta\tau^{DID}$	0.003	0.078	0.004	0.061	-0.005	0.015	0.053
	(0.023)	(0.042)	(0.032)	(0.048)	(0.024)	(0.026)	(0.035)
Adjustment term							
$\Delta\Psi$	-0.058	-0.051	-0.066	-0.051	-0.060	-0.058	-0.059
	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
Observations	361,091	361,091	361,091	361,091	361,091	361,091	361,091

Notes: Table A4 provides estimates of $\log ATET$ for the number of employees in different employee-categories using specifications (**??**) and (**??**). The table also provides estimates of the different components of these $\log ATET$ estimates. The estimates compare pre (2005-2007) and post (2011-2013) reform outcomes below and above the threshold in the treated group (firms with depreciation more than 10% of their labor costs) to the same differences in the control group (firms with depreciation less than 10% of their labor costs). We exclude the closest bin below and above the threshold to reduce the effect of bunching on the estimates. The estimates are provided for two different samples: the full sample is the same sample as shown in the bin-level graphs (i.e. firms with depreciation between 4.5k€ and 450k€), and another sample including firms with depreciation between 12.5k€ and 125k€.

Table A4: Reduced-Form Results with an Additional Control Group, Employee Categories

	(1)	(2)	(3)	(4)
	Total Investment	Investment in Fixed Assets	Investment in Buildings	Investment in R&D
Overall change with full sample				
$\log ATET^{es}$	0.181	0.211	0.188	0.094
	(0.042)	(0.039)	(0.162)	(0.064)
$\log ATET^{ns}$	0.128	0.151	0.152	0.057
	(0.040)	(0.036)	(0.162)	(0.063)
Log no. firms change with full sample				
$\Delta\tau_n^{DID}$	0.216	0.216	0.216	0.216
	(0.023)	(0.023)	(0.023)	(0.023)
Log mean change with full sample				
$\Delta\tau^{DID}$	-0.035	-0.005	-0.028	-0.122
	(0.035)	(0.029)	(0.161)	(0.059)
Adjustment term				
$\Delta\Psi$	-0.053	-0.060	-0.036	-0.037
	(0.005)	(0.005)	(0.003)	(0.003)
Observations	413,771	413,771	413,771	413,771

Notes: Table A5 provides estimates of $\log ATET$ for different investment categories using specifications (**??**) and (**??**). The table also provides estimates of the different components of these $\log ATET$ estimates. The estimates compare pre (2005-2007) and post (2011-2013) reform outcomes below and above the threshold in the treated group (firms with depreciation more than 10% of their labor costs) to the same differences in the control group (firms with depreciation less than 10% of their labor costs). We exclude the closest bin below and above the threshold to reduce the effect of bunching on the estimates. The estimates are provided for two different samples: the full sample is the same sample as shown in the bin-level graphs (i.e. firms with depreciation between 4.5k€ and 450k€), and another sample including firms with depreciation between 12.5k€ and 125k€.

Table A5: Reduced-Form Results with an Additional Control Group, Investment Categories

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 (**??**).

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

$$\partial_t f(S_t) = \partial_S g(S_t) f(S_t) - \frac{1}{2}\partial_{SS}\left[\sigma^2(S_t) f(S_t)\right], \tag{33}$$

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 (33) 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[14].

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

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

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

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

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

[14] 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 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

$$\tilde{f}_0(S) = L f_0(S).$$

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: $D = q f_1(\bar{S} - \delta) = L\tilde{f}_0(\bar{S} - \delta)$. 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 $l h_0(\bar{S})$ without the threshold, and $l h_1(\bar{S})$ with the threshold, as l is not affected by the slow growth interval. This implies that the probability of a firm that enters the slow growth band from below to exit above both with the threshold and without the threshold in place (i.e. $1 - p$) is equal to

$$1 - p = \frac{l f_1(\bar{S})}{l\tilde{f}_0(\bar{S})} = \frac{\exp\left(\frac{2\gamma\delta}{\sigma^2}\right) f_0(\bar{S})}{f_0(\bar{S})} = \exp\left(\frac{2\gamma\delta}{\sigma^2}\right).$$

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 – Derivations for Section 3

This appendix derives the results in Section 3. We use potential outcome notation, so that $y_t^g(R) = E(y_f(R)|f \in g \text{ during } t)$ 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 $R = 1$. The actual treatment group consists of three groups of firms: the above-stayers, the notch-switchers and the exit-switchers. This means that $\log ATET$ in the post-reform period $t = 1$ can be expressed as

$$\begin{aligned} \log ATET &= \log(E(y_{f,1}(1)|f \text{ is treated})) - \log(E(y_{f,1}(0)|f \text{ is treated})) \\ &= \log(E(y_1^g(1)|g \in \{as, ns, es\}))\log(E(y_1^g(0)|g \in \{as, ns, es\})). \end{aligned} \tag{34}$$

Assumptions 1-4 made in Section 3 are re-stated here for convenience:

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

$$\log(n_1^{bs}) - \log(n_0^{bs}) = \log(n_1^{as}) - \log(n_0^{as}) = \log(n_1^{ns}) - \log(n_0^{ns}) = \log(n_1^{es}) - \log(n_0^{es})$$

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

$$\log(y_1^{bs}(0)) - \log(y_0^{bs}(0)) = \log(y_1^{as}(0)) - \log(y_0^{as}(0)) = \log(y_1^{ns}(0)) - \log(y_0^{ns}(0))$$

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

$$\begin{aligned} y_t^{ns}(0) &= y_t^{bs} \\ y_t^{es}(0) &= 0 \end{aligned}$$

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

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

We observe the actual number of firms $n_{a,t}$ above ($a = 1$) and below ($a = 0$) the notch in each period and the actual average outcome $y_{a,t}$. From the definitions of the four groups

(i.e. from where they are located before and after the reform), we then have that

$$\begin{cases} n_{0,0} & = n_0^{bs} + n_0^{ns} \\ n_{1,0} & = n_0^{as} \\ n_{0,1} & = n_1^{bs} \\ n_{1,1} & = n_1^{as} + n_1^{ns} + n_1^{es} \end{cases} \tag{35}$$

and

$$\begin{cases} y_{0,0} & = y_0^{bs} \frac{n_0^{bs}}{n_0^{bs}+n_0^{ns}} + y_0^{ns} \frac{n_0^{ns}}{n_0^{bs}+n_0^{ns}} \\ y_{1,0} & = y_0^{as} \\ y_{0,1} & = y_1^{bs} \\ y_{1,1} & = y_1^{as} \frac{n_1^{as}}{n_1^{as}+n_1^{ns}+n_1^{es}} + y_1^{ns} \frac{n_1^{ns}}{n_1^{as}+n_1^{ns}+n_1^{es}} + y_1^{es} \frac{n_1^{es}}{n_1^{as}+n_1^{ns}+n_1^{es}}. \end{cases}$$

or

$$\begin{cases} y_{0,0} & = y_0^{bs}(0) \\ y_{1,0} & = y_0^{as}(0) \\ y_{0,1} & = y_1^{bs}(1) = y_1^{bs}(0) \\ y_{1,1} & = y_1^{as}(1) \frac{n_1^{as}}{n_1^{as}+n_1^{ns}+n_1^{es}} + y_1^{ns}(1) \frac{n_1^{ns}}{n_1^{as}+n_1^{ns}+n_1^{es}} + y_1^{es}(1) \frac{n_1^{es}}{n_1^{as}+n_1^{ns}+n_1^{es}}, \end{cases} \tag{36}$$

where we have introduced potential outcomes, used assumption 3 for the notch-switchers, and noted that as the below stayers are assumed not have been affected by the reform, $y_1^{bs}(1) = y_1^{bs}(0)$.

There are three treated groups: the above-stayers, the notch-switchers and the exit-switchers. Moreover, the definitions of the groups 1 imply that we assume that the treatment does not affect the number of firms in each group, but only where they are located. This is not very restrictive, as the exit-switchers are "located" in non-existence when the reform is not in place. The log average treatment effect of the treated is then given by

$$\begin{aligned} \log ATET = & \log \left(y_1^{as}(1) \frac{n_1^{as}}{n_1^{as}+n_1^{ns}+n_1^{es}} + y_1^{ns}(1) \frac{n_1^{ns}}{n_1^{as}+n_1^{ns}+n_1^{es}} + y_1^{es}(1) \frac{n_1^{es}}{n_1^{as}+n_1^{ns}+n_1^{es}} \right) \\ & - \log \left(y_1^{as}(0) \frac{n_1^{as}}{n_1^{as}+n_1^{ns}+n_1^{es}} + y_1^{ns}(0) \frac{n_1^{ns}}{n_1^{as}+n_1^{ns}+n_1^{es}} + y_1^{es}(0) \frac{n_1^{es}}{n_1^{as}+n_1^{ns}+n_1^{es}} \right) \end{aligned} \tag{37}$$

Note the expression for $y_{1,1}$ from 36 and that by assumption 3 ($y_1^{es}(0) = 0$) the expression

for $\log ATET$ in (37) can be expressed as

$$\log ATET = \log\left(y_{1,1}\right) - \log\left(y_1^{as}(0)\frac{n_1^{as}}{n_1^{as}+n_1^{ns}+n_1^{es}} + y_1^{ns}(0)\frac{n_1^{ns}}{n_1^{as}+n_1^{ns}+n_1^{es}}\right).$$

Moreover, using information from 35, we have

$$\log ATET = \log\left(y_{1,1}\right) - \log\left(y_1^{as}(0)\frac{n_1^{as}}{n_{1,1}} + y_1^{ns}(0)\frac{n_1^{ns}}{n_{1,1}}\right).$$

Now, the parallel trends assumption (assumption 2) implies that $y_1^{as}(0)/y_0^{as}(0) = y_1^{ns}(0)/y_0^{ns}(0) = y_1^{bs}(0)/y_0^{bs}(0)$. Combining this with 36, we have that all these are equal to $y_{0,1}/y_{0,0}$ as in the pre-period, the mean outcome below the threshold is equal to the mean outcome of below-stayer (as notch-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. Hence, $y_{0,1}/y_{0,0} = y_1^{bs}(0)/y_0^{bs}(0) = y_1^{as}(0)/y_0^{as}(0)$.

Therefore, we can express $\log ATET$ as

$$\begin{aligned}
\log ATET &= \log\left(y_{1,1}\right) - \log\left(\frac{y_1^{as}(0)}{y_0^{as}(0)}y_0^{as}(0)\frac{n_1^{as}}{n_{1,1}} + \frac{y_1^{ns}(0)}{y_0^{ns}(0)}y_0^{ns}(0)\frac{n_1^{ns}}{n_{1,1}}\right) \\
&= \log\left(y_{1,1}\right) - \log\left(\frac{y_1^{as}(0)}{y_0^{as}(0)}\right) - \log\left(y_0^{as}(0)\frac{n_1^{as}}{n_{1,1}} + y_0^{ns}(0)\frac{n_1^{ns}}{n_{1,1}}\right) \\
&= \log\left(y_{1,1}\right) - \log\left(\frac{y_{0,1}}{y_{0,0}}\right) - \log\left(y_0^{as}(0)\frac{n_1^{as}}{n_{1,1}} + y_0^{ns}(0)\frac{n_1^{ns}}{n_{1,1}}\right) \\
&= \log\left(y_{1,1}\right) - \log(y_{0,1}) + \log(y_{0,0}) - \log\left(y_{1,0}\frac{n_1^{as}}{n_{1,1}} + y_{0,0}\frac{n_1^{ns}}{n_{1,1}}\right) \\
&= \log\left(y_{1,1}\right) - \log(y_{1,0}) - \log(y_{0,1}) + \log(y_{0,0}) - \log\left(\frac{n_1^{as}}{n_{1,1}} + \frac{y_{0,0}}{y_{1,0}}\frac{n_1^{ns}}{n_{1,1}}\right),
\end{aligned}$$

where the second equality follows from the parallel trends assumption for means (assumption 2), the third equality follows from the previous paragraph, the fourth equality follows from Equation 36 and assumption 3.

Hence, we have that

$$\log ATET = \underbrace{\tau^{DID}}_{\text{Diff-in-diff estimate}} \underbrace{- \log\left(\frac{n_1^{as}}{n_{1,1}} + \frac{y_{0,0}}{y_{1,0}}\frac{n_1^{ns}}{n_{1,1}}\right)}_{\text{Switcher term}}, \tag{38}$$

which is the same as Equation (16).

C.1 Only exit-switchers

Here, we show that $\log ATET$ is point-identified under the assumption that all switchers are exit-switchers. The corresponding assumption is:

Assumption 5 (No notch-switchers.)**.**

$$n_t^{ns} = 0$$

Under assumption 5, the expression for $\log ATET$ in 38 becomes.

$$\log ATET = \tau^{DID} - \log\left(\frac{n_1^{as}}{n_{1,1}}\right).$$

Moreover, combining assumption 5 with Equation 35, we have that $n_{0,0} = n_0^{bs}$. This together with the assumption of parallel trends in log number of firms (assumption 1) implies that $n_1^{as}/n_0^{as} = n_1^{bs}/n_0^{bs} = n_{0,1}/n_{0,0}$. Using this and $n_0^{as} = n_{1,0}$ from 35, we have

$$\begin{aligned}\log ATET &= \tau^{DID} - \log\left(\frac{n_1^{as}}{n_0^{as}}\frac{n_0^{as}}{n_{1,1}}\right) \\ &= \tau^{DID} - \log\left(\frac{n_{0,1}}{n_{0,0}}\frac{n_{1,0}}{n_{1,1}}\right)\end{aligned}$$

so that

$$\log ATET = \tau^{DID} + \tau_n^{DID},$$

where $\tau_n^{DID} = \log(n_{1,1}) - \log(n_{1,0}) - (\log(n_{0,1}) - \log(n_{0,0}))$ is the difference-in-differences estimate of log number of firms comparing pre- and post-reform periods above and below the notch.

C.2 Only notch-switchers

Here, we show that $\log ATET$ is identified under the assumption that all switchers are notch-switchers. The corresponding assumption is:

Assumption 6 (No exit-switchers.)**.**

$$n_t^{es} = 0$$

Combining assumption 6 with Equation 35, we have that $n_1^{ns} = n_{1,1} - n_1^{as}$, so that the

expression for $\log ATET$ in Equation 38 becomes

$$\log ATET = \tau^{DID} - \log\left(\frac{n_1^{as}}{n_{1,1}} + \frac{y_{0,0}}{y_{1,0}}\left(1 - \frac{n_1^{as}}{n_{1,1}}\right)\right).$$

Moreover, assumption 6 together with the assumption on parallel log trends in the number of firms of each group (assumption 1) the total – and observable – number of firms in the economy $n_t = n_t^{as} + n_t^{bs} + n_t^{ns}$ has the same trend as all sub-groups $g \in \{bs, as, ns\}$ as there are no firms that enter/exit due to the reform. Hence, we have that $n_1^{as}/n_0^{as} = n_1/n_0$. Also, note from 35 that $n_0^{as} = n_{1,0}$. With these, the expression for $\log ATET$ becomes

$$\log ATET = \tau^{DID} - \log\left(\frac{n_1}{n_0}\frac{n_{1,0}}{n_{1,1}} + \frac{y_{0,0}}{y_{1,0}}\left(1 - \frac{n_1}{n_0}\frac{n_{1,0}}{n_{1,1}}\right)\right)$$

or

$$\log ATET = \tau^{DID} + \log(n_{1,1}) - \log(n_{1,0}) - (\log(n_1) - \log(n_0)) - \log\left(1 + \frac{y_{0,0}}{y_{1,0}}\left(\frac{n_{1,1}}{n_{1,0}}/\frac{n_1}{n_0} - 1\right)\right).$$

Adding and subtracting $\log(n_{0,1}) - \log(n_{0,0})$, this becomes

$$\log ATET = \tau^{DID} + \tau_n^{DID} + \log\left(\frac{n_{0,1}}{n_{0,0}}/\frac{n_1}{n_0}\right) - \log\left(1 + \frac{y_{0,0}}{y_{1,0}}\left(\frac{n_{1,1}}{n_{1,0}}/\frac{n_1}{n_0} - 1\right)\right).$$

Defining $f_{a,t} \equiv n_{a,t}/n_t$ as the fraction of firms above ($a = 1$) or below ($a = 0$) the threshold in period t, we can express this as

$$\begin{aligned}\log ATET &= \tau^{DID} + \tau_n^{DID} + \log\left(\frac{f_{0,1}}{f_{0,0}}\right) - \log\left(1 + \frac{y_{0,0}}{y_{1,0}}\left(\frac{f_{1,1}}{f_{1,0}} - 1\right)\right) \\ &= \tau^{DID} + \tau_n^{DID} + \Psi,\end{aligned}$$

which is the same as in Equation (20).