Firms' Responses to Tax Enforcement Strategies: Evidence from Spain

Almunia, Miguel and Lopez-Rodriguez, David

University of Warwick, Banco de España

15 November 2013

Firms' Responses to Tax Enforcement Strategies: Evidence from Spain

Miguel Almunia & David Lopez-Rodriguez*

November 15, 2013

Abstract

We study the effects of tax enforcement policies on firm behavior in Spain. Firms with more than €6 million in reported revenue are monitored by the Large Taxpayers' Unit (LTU), which results in more frequent tax audits and more information requirements for those firms. We exploit this discontinuity in tax enforcement intensity to estimate the impact of different enforcement regimes on firms' reporting behavior, using financial statements from almost the universe of Spanish firms for the period 1999-2007. An excess mass of firms locates, or "bunches", just below the revenue threshold. Using bunching estimation techniques, we calculate that the marginal firm reduces its reported revenue by up to 7.5% to avoid falling in the high enforcement regime. A dynamic analysis shows that firm's revenue growth rates decline substantially as firms approach the LTU threshold from below. We provide suggestive evidence that firms locating below the threshold also misreport their material and labor input costs to evade taxes.

Keywords: tax enforcement, firms, bunching, Spain, Large Taxpayers Unit (LTU).
JEL codes: H26, H32, K42.

---

*Almunia: University of Warwick Department of Economics and CAGE; M.Almunia@warwick.ac.uk. Lopez-Rodriguez: Banco de España; david.lopezr@bde.es. We gratefully acknowledge useful comments and suggestions from Emmanuel Saez, Alan Auerbach, Fred Finan, Ted Miguel, Juan Pablo Atal, Michael Best, David Card, Lorenzo Casaburi, Raj Chetty, François Gerard, Jonas Hjort, Simon Jäger, Attila Lindner, Justin McCrary, Craig McIntosh, Adair Morse, Gautam Rao, Ana Rocca, Michel Serafinelli, Monica Singhal, Juan Carlos Suárez Serrato, Victoria Vanasco, Andrea Weber, Danny Yagan, Owen Zidar, and numerous seminar participants at UC Berkeley, Universidad de Barcelona, La Caixa, EEA 2012 Meetings, CIDE, Banco de México, Fundaçao Getulio Vargas, U. of Illinois at Urbana-Champaign, Mannheim, U. Autònoma de Barcelona, IESE, Warwick, Instituto de Empresa, CEMFI, and Oxford CBT. Almunia gratefully acknowledges financial support from Fundación Rafael del Pino and the Burch Center for Tax Policy and Public Finance.

1 Introduction

Firms remit more than three-quarters of the tax revenues raised by governments in advanced economies.[1] Despite playing such a crucial role as fiscal intermediaries, the empirical literature on tax evasion has largely neglected firms, focusing instead on individual cheating behavior.[2] Absent perfect tax monitoring, the information asymmetry between businesses and tax authorities gives firms incentives to misreport their own income, and also third party's income, in order to evade taxes.

This paper takes advantage of a policy discontinuity to analyze how tax enforcement policies (i.e., tax audits and compliance requirements) affect firms' reporting and production decisions. In 1995, the Spanish tax agency established a Large Taxpayers' Unit (LTU) to monitor and enforce the taxation of companies with annual operating revenues above €6 million.[3] Firms assigned to the LTU are subject to more frequent tax audits and information cross-checking by the tax authority, while the tax schedules is the same on either side of the threshold. This discontinuity in tax enforcement intensity gives tax-evading firms an incentive to remain below the revenue threshold. They can do this either by reducing their output or by underreporting their revenue or a combination of both reactions. In this paper, we first estimate the size of the total reported revenue response and then we examine whether firms' reaction is due to an adjustment in real production or to revenue misreporting, which implies tax evasion.

To guide the empirical estimation, we set up a theoretical framework where profit maximizing firms decide (i) how much to produce and (ii) how much of their revenue to underreport in order to reduce their tax liability. Firms receive a productivity draw that determines their optimal size in equilibrium. The available enforcement technology available yields a probability of evasion detection that is continuously increasing in firm size and in the amount evaded. This reflects the intuition that larger firms are more visible to the tax authority, and that egregious evasion is easily detectable. We introduce the concept of a LTU in the model by allowing the detection probability to jump up discretely at a fixed level of reported revenue. This generates a notch in tax enforcement, meaning that the probability of detection increases discontinuously when a firm crosses the threshold.[4] The existence of the notch

[1]In the United States, for example, businesses remit 84 percent of taxes (Christensen et al., 2001).

[2]Slemrod (2004, 2008) has repeatedly stressed the relevance of firms in the analysis of tax evasion.

[3]Firms in the LTU represent only 2.5% of all registered business, but they employ 50% of private sector workers and report 75% of taxable profits (AEAT, Several years). Most tax agencies in advanced countries, and an increasing number of emerging countries, have some type of LTU to deal with large businesses (see IMF, 2002 and OECD, 2011).

[4]See Slemrod (2010) for a formal definition of notches in tax schedules and regulations.

drives some firms to report lower revenue and bunch at the LTU threshold to avoid high tax enforcement. We define the "marginal buncher" as the firm with the highest productivity draw that chooses to bunch at the cutoff point.

In the empirical analysis, we use financial statements and balance-sheet data from Amadeus. This database compiles information reported by firms to the Commercial Registry, covering more than 80% of registered businesses in Spain with operating revenue between €3 and €10 million for the period 1999-2011 (see Table 6 for details). The longitudinal structure of the dataset allows us to analyze the dynamic behavior of firms. An advantage of this data source over administrative tax returns is that it provides information on firms' choices in several dimensions, such as input costs, allowing us to examine the impact of multiple taxes in a single data source.[5]

To estimate the firms' response to a discontinuity in tax enforcement intensity, we analyze the empirical distribution of reported operating revenue in Spain. As predicted by the model, a significant number of firms bunch below the LTU threshold. This behavioral response is strong and persistent over time for the boom period (1999-2007), but becomes much smaller in the recession period (2008-2011). The evidence indicates that the bunching response is due exclusively to the existence of the LTU, not to other regulations affecting firms in the same size range such as tax incentives for small firms.

We construct a counterfactual revenue distribution and use it to estimate the excess bunching mass in a short interval below the threshold. We then use this excess mass as a sufficient statistic for the reported revenue response of the marginal buncher. Despite the notch in enforcement, many firms report revenues just above the LTU threshold, suggesting the existence of optimization frictions. We use the missing mass in an interval above the threshold (where the bunching firms would have located in the absence of the policy) as a proxy for the degree of optimization frictions. Dividing the original bunching estimator by this proxy, we obtain a treatment-on-the-treated estimator of the total revenue response.

For the boom period 1999-2007, we estimate that the marginal buncher reduces its reported revenue by €86,000 (1.4% of total reported revenue) under the assumption of no optimization frictions, and €449,000 (7.5% of total reported revenue) once frictions are taken into account. This is a sizable response, considering that average reported profits around the LTU threshold are €290,000 (4.5% of revenue). The estimates are significantly

[5] In order to perform our empirical tests using administrative data, we would need tax returns from the corporate income tax, the value added tax, and social security contributions. It is rare for researchers to have access to all these sources of information simultaneously, and especially to be able to link them (since administrative datasets are often anonymized).

smaller for the recession period 2008-2011, where the "no frictions" estimate is €26,000 and the "frictions" estimate is €384,000.

There is heterogeneity in the bunching response across different groups of firms. Bunching is somewhat stronger among firms that are small in non-revenue dimensions such as fixed assets or number of employees. Across sectors of activity, there is an inverted-U relationship between the size of the response in a given sector and a "scope of evasion" index that takes into account the median number of employees and the share of output sold to final consumers in each sector. There is also wide regional variation, with the strongest bunching in the Central and Southern regions and the smallest in the North-East.

To complement our static bunching analysis we also analyze the dynamic behavior of firms. Growing firms, defined as those reporting higher revenue in the current year than last year, are much more likely to bunch at the threshold than shrinking firms, which barely respond. The probability of revenue growth and the median growth rate drop significantly as firms approach the threshold from below. Moreover, the probability to remain in the same €250,000-wide revenue bin for two consecutive years almost doubles for firms just below the threshold compared to the counterfactual.

To better understand whether firms' reaction to the tax enforcement discontinuity is due to real or evasion responses, we extend the stylized model of firm behavior to include two production inputs, labor and materials, that can be misreported. Each type of input leads to different fiscal incentives: materials are deductible under both the corporate income tax (CIT) and the value added tax (VAT), so it is always advantageous to overreport them to evade taxes. In contrast, labor costs are not deductible under the VAT but they are taxed through the payroll tax (PRT). The incentives for a tax-evading firm regarding labor costs are therefore ambiguous from this perspective, depending on the marginal tax rates on the CIT and the PRT. However, the existence of downward nominal wage rigidities and uncertainty about future demand conditions make it reasonable to assume that a tax-evading firm has incentives to underreport labor costs.

We use this model extension to derive testable predictions about the behavior of reported input costs around the LTU threshold. When bunching is due to a real response, the model predicts an upward jump in reported input costs at the threshold (that is, firms just below the threshold report less labor and material inputs than firms just above). This follows from the fact that bunching firms have higher productivity and the production function is concave, so bunching firms need fewer inputs to produce the same output as other firms. When bunching is due to tax evasion, the model predicts a downward jump in reported material inputs at the LTU threshold. This is due to bunching firms producing more output (and hence using more inputs) than they report. Since we expect bunching firms to also overreport their material inputs, the jump will be even more pronounced. In contrast, the model predicts a downward jump in reported labor costs, for exactly the opposite reasons.

We compare the predictions of the model with empirical patterns by showing plots of the ratio of reported inputs costs (materials, labor) over revenue against reported revenue. Using ratios rather than levels helps to identify discontinuities at the LTU threshold. In the boom period, we find a significant downward jump in average reported material inputs (as a percentage of revenue) at the threshold, from about 66% to 64.5%, meaning that firms just below report proportionally more material inputs than firms just above. Such a pattern is at odds with the predictions of the model where bunching is due to real responses. We also find an upward jump in reported labor costs at the threshold, from 15% to 16% of revenue. Most of this discontinuity in reported labor expenses is due to firms below the threshold reporting lower average wages than firms above, while they report a similar number of employees. The patterns observed for reported labor inputs seems to point to evasion responses, whereby bunching firms underreport their employees' wages (to evade payroll taxes) so that their reported labor costs are lower than those of firms above the threshold. Overall, the graphical evidence rules out the hypothesis that bunching is due entirely to a real response. However, the evidence is not enough to prove that it is all due to evasion, nor can we disentangle the importance of revenue vs. input cost misreporting.

The findings in this paper contribute to the scarce empirical literature on business tax evasion by providing a well-identified measure of the effects of tax enforcement on firm behavior. Recent research by De Paula and Scheinkman (2010) and Pomeranz (2013) emphasizes the role of information for effective tax enforcement, particularly in the presence of self-enforcing mechanisms related to the VAT.[6] While Pomeranz (2013) concludes that having a paper trail of transactions is complementary to tax audits, our results suggest that audits are necessary for paper trails to be useful in tax enforcement.

The empirical techniques used below draw on a recent literature in public finance that analyzes agents' responses to thresholds in taxes and regulations. The seminal paper is Saez (2010), but our empirical strategy draws mostly on the work of Kleven and Waseem (2013), who exploit income thresholds at which the average tax rate jumps to identify taxable income elasticities.[7] The novel feature of our setting within this literature is that the LTU generates

[6]In fiscal systems with a VAT, the transmission of evasion (or compliance) behavior moves upwards the production chain from retailers to intermediate goods suppliers.

[7]Saez (2010) exploits kink points in the US personal income tax schedule – i.e., income thresholds at

a notch in the probability of evasion detection, rather than the tax rate (which is unaffected in this setting), allowing us to study the effect of tax enforcement in isolation.

This paper is also related to the recurring challenge in tax enforcement policy of how to effectively monitor small businesses. Evasion becomes riskier and more costly to firms as they get bigger, because they need sophisticated accounting systems to carry out complex operations. This facilitates the tax agency's task of obtaining information from large firms[8] (Kleven et al., 2009). Such information-related constraints on tax evasion are much weaker in the case of small businesses, which represent the vast majority of firms.[9] This is particularly relevant in the Spanish case, where average firm size is small given the country's level of development.[10] Since the expected return from a tax audit grows more than proportionally with firm size, Dharmapala et al. (2011) make the theoretical argument that it may be optimal for tax agencies with limited resources to focus all their enforcement efforts on large firms.

Finally, our paper contributes to the extensive literature on firm size distribution[11] and size-contingent policies.[12] This literature focuses on the productivity consequences of having a distorted firm size distribution with too many small firms. One of our contributions is to point out that some firms may look smaller in the data than they actually are due to misreporting. A closely-related paper is Onji (2009), who studies how Japanese responded to the introduction of a VAT by trying to remain under the VAT eligibility threshold.

The rest of the paper is organized as follows. Section 2 presents the theoretical framework to guide our empirical analysis. Section 3 provides some institutional background on tax enforcement policies in Spain and describes the dataset. Section 4 presents the static empirical analysis, including a derivation of the bunching parameters. Section 5 presents the dynamic empirical analysis. Section 6 explores real and evasion components of the response of firms. Section 7 discusses the welfare implications of the results. Section 8 concludes.

---

which the marginal tax rate jumps – to identify taxable income elasticities. Several recent studies Chetty et al. (2011, 2012); Bastani and Selin (2012) apply Saez's method to derive taxable income elasticities using large administrative datasets from Denmark, Sweden and the United States. Devereux et al. (2013) also use bunching techniques to estimate the elasticity of corporate taxable income in the United Kingdom.

[8]On the other hand, large firms tend to spend more resources to hire top accountants and lawyers to maximize legal tax *avoidance*.

[9]The exact percentage depends on the country and the precise definition of what constitutes a small firm.

[10]The share of small firms seems to be positively correlated with the size of the underground economy across countries. Schneider et al. (2010) estimate that the underground economy accounts for approximately 25% of GDP in Greece, Italy and Spain, where the firm size distribution is skewed towards small family firms. This is high compared to about 15% in France and Germany, and less than 10% in the United States, where firms are larger on average.

[11]To name just a couple, Lucas (1978), Cabral and Mata (2003).

[12]Some examples are **?** and Restuccia and Rogerson (2008).

2 Theoretical Framework

We model the problem of a profit-maximizing firm that can choose to evade part of its tax liabilities, at the risk of paying a penalty if it gets caught. The basic setup extends the classic individual tax evasion framework (e.g., Allingham and Sandmo, 1972) to firms. We enrich this framework by endogenizing the probability of detection, making it depend on firm size and on the amount of evasion.

2.1 Setup

Consider an economy with a continuum of firms of measure one. Firms produce good y using inputs m according to the production function $y = \psi f(m)$, where ψ is an idiosyncratic productivity parameter distributed over the range $[\underline{\psi}, \overline{\psi}]$ with probability density function (pdf) $h_0(\psi)$. The production function exhibits positive but decreasing returns to material inputs ($f_m > 0, f_{mm} < 0$). All markets are competitive, so firms purchase inputs at price c and sell all their output at price p (which we normalize to 1 for simplicity). There is no entry or exit of firms, such that in equilibrium all firms with $\psi > \underline{\psi}$ can sustain positive profits.

The government sets a proportional tax t on profits, so after-tax profits are given by $\Pi = (1-t)\left[\psi f(m) - cm\right]$. Assuming that tax evasion is not possible, profit maximization yields the standard condition:

$$\psi f_m\left(m^{NoEv}\right) = c \tag{1}$$

where m^{NoEv} is the optimal input use when there is no evasion. Given the definition of y, this defines optimal true production $y^{NoEv} = \psi_i f\left(m^{NoEv}\right)$. The proportional tax on profits does not distort production efficiency in this simple partial equilibrium setting. Firms optimize production as they would without taxation, but they now transfer part of their profits to the government.

Now assume that firms can evade taxes by underreporting their revenue, which reduces their tax liability. Let $u \equiv y - \overline{y}$ denote the amount of revenue underreported, where $\overline{y}$ is reported revenue. We assume that input costs are always reported truthfully, so reported profits are given by $\overline{\Pi} = (1-t)\left[\overline{y} - cm\right]$. The tax agency detects tax evasion with probability $\delta \in (0,1)$, which is endogenously determined as we explain below. We think of δ as the audit probability, and we make the simplifying assumption that evasion is always detected if there is an audit. When evasion is detected, a penalty rate θ is applied on the total amount evaded, and after-tax profits are given by $\Pi^D = (1-t)\Pi - \theta t[\Pi - \overline{\Pi}]$. If no evasion is detected, after-tax profits are $\Pi^{ND} = \Pi - t\overline{\Pi}$.

We can then write expected after-tax profits as follows:

$$\mathbb{E}\Pi = (1-\delta)\,\Pi^{ND} + \delta\Pi^{D}$$
$$= (1-t)\,\Pi + tu\left[1-\delta\,(1+\theta)\right]. \tag{2}$$

2.2 Benchmark Case

Let the probability of detection $\delta = \delta\,(u, m)$ be a continuous and strictly monotonic function of evasion and true input use. We assume that $\delta_m\,(u, m) > 0$ (which implies $\delta_y\,(u, y) > 0$ because the production function is monotonically increasing), to capture the intuition that larger firms are more visible and hence more likely to be audited by the tax agency.[13] Additionally, we assume that $\delta_u\,(u, m) > 0$, which has two important implications. First, firms face a trade-off between the benefits of evasion (lower tax payments) and the increased probability of detection. Second, the tax agency's enforcement strategy is influenced by the reporting behavior of firms. One way to motivate this assumption is to consider commonly used "relative audit rules", under which tax agencies use aggregate information obtained from firms in similar markets to identify suspicious behavior (Bayer and Cowell, 2009). For example, a company operating in a booming industry that reports negative profits is very likely to be audited because it stands out from its peers.[14]

The probability of detection is common knowledge. To ensure that the probability is bounded, we further assume that $\lim_{u\to 0}\delta\,(u, y) = 0$ and $\lim_{u\to y}\delta\,(u, y) = \frac{1}{1+\theta}$. The latter condition implies that the detection technology is not perfect, because even when a firm reports zero revenue there is no certainty that it will be detected. This assumption is also convenient to rule out corner solutions and ensure that all firms have a positive amount of underreporting in equilibrium. We assume that $\delta\,(u, m)$ is locally convex in the neighborhood of y^{LTU}, i.e. $\delta_{uu}\left(u, m|\bar{y} \approx y^{LTU}\right) > 0$.

Firms simultaneously make production (m) and reporting (u) decisions to maximize expected profits. Optimal conditions for an interior optimum are given by:

$$\psi f_m\,(m^*) = c + u\left[\frac{t}{1-t}\right][1+\theta]\,\delta_m\,(u, m^*) \tag{3}$$

$$1 = [1+\theta]\left[\delta\,(u^*, m) + u^*\cdot\delta_u\,(u^*, m)\right] \tag{4}$$

[13] The idea of an endogenous probability of detection that depends positively on the amount evaded was first introduced by Reinganum and Wilde (1985).

[14] Notice that this type of audit rule provides "good" incentives, because firms are better off reporting higher profits in order to avoid tax audits, holding all else equal.

Condition (3) is similar to the standard optimality condition (1), but with an additional positive term on the right-hand side. This term accounts for the fact that higher production increases the probability of detection. Since $u \geq 0$ by definition, in an interior optimum we obtain that $m^* < m^{NoEv}$, which implies $y^* < y^{NoEv}$. In the corner solution where $u^* = 0$, condition (3) reduces to (1). Comparative statics are intuitive: optimal input use m^* is larger when (i) its effect on the detection probability is weaker (i.e., $\delta_m(u, m)$ is smaller), (ii) the tax rate t or the penalty θ are lower, and (iii) the equilibrium amount of underreported revenue u^* is smaller.

Condition (4) equates the expected marginal benefit of an additional unit of evasion to the expected marginal cost. Firms optimally choose to underreport sales as long as $\delta(1 + \theta) < 1$, which we assumed above. Comparative statics show that optimal evasion u^* is higher when (i) the penalty rate θ is lower, and (ii) the probability of detection δ is lower.[15]

The analysis above shows that, when enforcement policies respond endogenously to firms' production and reporting decisions, such policies will in turn affect firm behavior. Compared to the situation with no evasion, firms produce less output and engage in revenue underreporting. These results are qualitatively similar to those obtained by Bayer and Cowell (2009) in a model where they explicitly introduce relative audit rules. Since the production and cost functions are the same for all firms, each firm's optimal size in equilibrium depends uniquely on their idiosyncratic productivity level ψ. It can be shown that if the productivity distribution $h_0(\psi)$ is smoothly decreasing in its full domain $\left[\underline{\psi}, \overline{\psi}\right]$, then there exists a density function $g_0(\cdot)$ such that the distribution of firms' operating revenue, $g_0(\overline{y}^*)$ is also smoothly decreasing in its full domain $\left[\overline{y}^*\left(\underline{\psi}\right), \overline{y}^*\left(\overline{\psi}\right)\right]$.[16]

[15] We apply the Implicit Function Theorem to do the comparative statics of an increase in the probability of detection. Let $F(u, \delta) \equiv \frac{d}{du}\mathbb{E}\Pi = 1 - [1 + \theta][\delta + u^* \cdot \delta_u(u^*, m)]$. Then:

$$\begin{aligned}
\frac{du}{d\delta}|_{u=u^*} &= -\frac{dF/d\delta}{dF/du}|_{u=u^*} \\
&= -\frac{1+\theta}{[1+\theta][\delta_u + \delta_u + u^*\delta_{uu}]}|_{u=u^*} \\
&= -\frac{1}{2\delta_u + u^*\delta_{uu}}|_{u=u^*} \\
&< 0, \text{ since } \delta_u, \delta_{uu} > 0.
\end{aligned}$$

[16] The specific mapping between the two density functions depends on the functional forms of the production function $f(m)$ and the probability of detection $\delta(u, m)$.

2.3 Policy Intervention: Large Taxpayers' Unit (LTU)

Assume now that the tax agency sets up a Large Taxpayers' Unit that monitors and enforces the taxation of firms with reported revenue higher than y^{LTU}. Dharmapala et al. (2011) provide a theoretical rationale for the existence of this type of institution when the tax agency's resources are limited. In their model, the trade-off between the tax agency's administrative costs of enforcement and its tax collection goals yields an optimal threshold below which firms should be exempted from taxation.[17] They argue that the full exemption for small businesses exists *de facto* in most developing countries via lenient tax enforcement.

The probability of detection is no longer a continuous function of reported revenue. It remains the same for firms below the revenue cutoff and jumps discretely at the revenue threshold y^{LTU}. Hence, the detection probability is strictly higher for all firms above the threshold and given by:

$$\delta = \begin{cases} \delta\left(u,m\right), & \text{if } \overline{y} \leq y^{LTU} \\ \delta^{LTU} \equiv r \cdot \delta\left(u,m\right), & \text{if } \overline{y} > y^{LTU} \end{cases}$$

where, $r > 1$. We assume that $\delta\left(\cdot\right)$ is locally convex at y^{LTU} such that the optimal conditions (3) and (4) continue to hold for firms with $\bar{y} \leq y^{LTU}$.

The introduction of the LTU generates a "notch" in δ, meaning that the probability of detection increases for *all* inframarginal units evaded when a firm crosses the (reported) revenue threshold. We assume that firms face no optimization frictions (we relax this assumption later), so they can re-optimize to new levels of production and reporting in response to the new policy. The pre-reform and post-reform revenue distributions are depicted in Figure 2, where they are labeled "counterfactual" and "observed" density, respectively, to be consistent with the terminology of the empirical section.

To study the response of different types of firms to the policy change, we define three distinct groups. First, there are low productivity firms, defined as those that report revenue $\overline{y}^* \leq y^{LTU}$ in the benchmark case. Nothing changes for these firms with the new policy because they are not LTU-eligible, so their behavior continues to be defined by optimality conditions (3) and (4). We denote by ψ^L the productivity level of the firm that chooses exactly $\overline{y}^* = y^{LTU}$ in the benchmark case (without LTU). Hence, all firms with $\psi_i \in \left[\underline{\psi}, \psi^L\right]$ belong to the "low productivity" group.

[17] The threshold in Dharmapala, Slemrod and Wilson (2011) involves changes in both tax liability and enforcement, whereas in our setting only the enforcement intensity changes.

Second, there is a group of firms whose pre-reform reported revenue was just above y^{LTU}. These firms react to the reform by reporting lower revenue in order to locate exactly, or "bunch", at the LTU threshold, i.e. $\bar{y}^{**} = y^{LTU}$ (we denote the optimal choices in the LTU case with two stars, to distinguish them from optimal choices in the benchmark case, which had one star). This bunching response is a combination of lower production and higher evasion, where the relative importance of each action depends on the functional forms of $f(m)$ and $\delta(u, m)$. We define the "marginal buncher" as the firm with the highest exogenous productivity that chooses $\overline{y}^{**} = y^{LTU}$. We denote by ψ^{MB} the exogenous productivity of the marginal buncher. Formally, ψ^{MB} is the unique value that equalizes expected profits when facing the low probability of detection (δ) an expected profits when facing the high probability $\left(\delta^{LTU}\right)$:

$$\mathbb{E}\Pi\left(u^{**}, m^{**}|\psi^{MB}, \delta\right) = \mathbb{E}\Pi\left(u^{**}, m^{**}|\psi^{MB}, \delta^{LTU}\right) \tag{5}$$

An important point to notice about expression (5) is that the optimal values (u^{**}, m^{**}) are different under each probability of detection. Given the above definitions, all firms with productivity $\psi \in \left(\psi^L, \psi^{MB}\right]$ belong to the group of "bunching firms".

Third, there is a group of high productivity firms, with $\psi > \psi^{MB}$, which are affected by the introduction of the LTU because they now face a higher probability of detection. For these firms, reducing reporting revenue all the way to y^{LTU} is too costly because it involves either inefficiently low production or too much exposure to being detected by the tax agency (or both). The optimality conditions faced by these firms are equivalent to (3) and (4), but with $\delta^{LTU}(u, m)$ instead of $\delta(u, m)$. Hence, these "high productivity" firms re-optimize and report higher revenue than they did in the benchmark case: $\overline{y}^{**}\left(\psi > \psi^{MB}\right) > \overline{y}^{*}\left(\psi > \psi^{MB}\right) > y^{LTU}$.

We can sum up the characterization of these three groups of firms as a function of exogenous productivity levels:

- If $\psi_i \in \left[\underline{\psi}, \psi^L\right]$, firm i is a Low Productivity Firm
- If $\psi_i \in \left(\psi^L, \psi^{MB}\right]$, firm i is a Bunching Firm
- If $\psi_i \in \left(\psi^{MB}, \overline{\psi}\right]$, firm i is a High Productivity Firm

Bunching firms are the most important group for our analysis. We use a first-order approximation to relate the number of bunching firms and the reported revenue response of the marginal buncher. For analytical simplicity, consider the case where the LTU raises the detection probability by an arbitrarily small amount $d\delta \equiv \delta^{LTU}(\cdot) - \delta(\cdot)$. In this case, the range of bunching firms would also be arbitrarily small and we can define $d\psi \equiv \psi^{MB} - \psi^{L}$, which is the difference in exogenous productivity between the marginal buncher and the largest of the low productivity firms. In the benchmark case, we established that there is a direct mapping from the pdf of the productivity parameter, $h_0(\psi)$, to the pdf of reported revenue, $g_0(\overline{y})$. Hence, we can define the excess mass of bunching firms, B, as follows:

$$B = \int_{y^{LTU}}^{y^{LTU}+d\overline{y}} g_0(\overline{y})\, d\overline{y} \approx g_0\left(y^{LTU}\right) d\overline{y}^{MB}, \tag{6}$$

where the approximation assumes that the counterfactual density $g_0(\overline{y})$ is approximately flat in the neighborhood of y^{LTU}. The term $g_0\left(y^{LTU}\right)$ denotes the height of the density distribution at the LTU threshold (in the benchmark case), while $d\overline{y}^{MB}$ is the change in reported revenue for the marginal buncher in response to the introduction of the LTU.[18] Under the strong assumption that firms face no optimization frictions[19], $d\overline{y}^{MB}$ can also be interpreted as the length (in million Euros) of the interval were the density is zero, as shown in Figure 2. To be able to estimate this amount, we use (6) to define the parameter b as the ratio of excess bunching over the counterfactual density at the threshold**:**

$$b \equiv \frac{B}{g_0\left(y^{LTU}\right)} \approx d\overline{y}^{MB} \tag{7}$$

In Section 4.1, we develop an empirical strategy to build a counterfactual distribution and calculate the excess bunching mass in order to estimate b in the data. We refer to parameter b as a measure of "bunching intensity". In Section 4.2, we relax the assumption of no optimization frictions and define an alternative estimator of b that takes frictions into account.

[18]In the benchmark scenario, the marginal buncher reported $\overline{y}_0^{MB} = y^{LYU} + d\overline{y}$, but in presence of the LTU this firm reports $\bar{y}^{**} = y^{LTU}$.

[19]We discuss at length the implications of the existence of optimization frictions in Section 4.2.

3 Institutional Background and Data

In the past three decades, multilateral institutions (IMF, 2002; OECD, 2011) have advised national governments to set up large taxpayer units (LTU) in order to increase control over the largest taxpayers and improve their tax compliance, based on the success of early adopters of this policy. We summarize below the main characteristics of the Spanish LTU, the main source of variation we exploit in this paper. We also describe two other policy thresholds that are relevant to our empirical analysis.

3.1 Tax Administration Thresholds

LTU threshold

The Spanish tax agency established a LTU (*Unidades Regionales de Gestión de Grandes Empresas*) in 1995, with an office in each of the 17 Spanish regions. The threshold to define a "large firm" was set at €6 million in annual operating revenue and has not been modified since then.[20] Exporters that claim reimbursements on their VAT return are always included in the LTU, regardless of their total reported revenue.[21] When a firm reports revenue above the threshold in a given year t, it is automatically added to a 'census' of large firms in year $t + 1$. Notice that the first annual tax return that the LTU can audit is that of year t, due to the timing of tax reporting. Generally, the law allows tax auditors to revise tax returns up to four years in the past.

Firms in the LTU census are subject to stricter monitoring and higher compliance requirements. The LTU performs comprehensive tax audits on approximately 10% of large firms each year, while barely 1% of firms below the threshold are audited (AEAT, Several years). Moreover, firms in the LTU are required to file their VAT declarations on a monthly basis (instead of quarterly) and in electronic form (as opposed to on paper). The latter makes it easier to cross-check individual transactions of LTU firms compared to firms below the threshold.[22] Finally, the withholding rate on the corporate income tax is 25%, compared to 18% for small firms, which could potentially create liquidity constraints for LTU firms.

[20] The threshold was originally set at 1 billion pesetas, the official currency at the time. The fixed exchange rate is 166.386 pesetas per euro, so the threshold is exactly at €6.010121 million. In 2006, an additional threshold of €100 million in operating revenue was established to determine eligibility to the Central Delegation for Large Firms, a select group of the largest firms within the LTU. We observe no evidence of bunching at this revenue threshold.

[21] VAT evasion through so-called "carousel fraud" is a first-order issue at the European Union level (see Keen and Smith (2007) for details).

[22] A recent reform extended electronic reporting to all firms since July 1st, 2008.

The number of firms in the LTU census increased from 18,860 (2.4% of all registered firms) in 1999 to 40,571 (2.9%) in 2007, due to economic growth and an annual inflation rate around 3%. In the same period, the overall staff of the tax agency remained almost constant, but enforcement intensity is likely to have been stable or even increased thanks to technological improvements. Firms in the LTU report about 80% of all taxable profits each year and employ more than 40% of private sector employees (AEAT, Several years).

Corporate Income Tax Threshold

The standard rate in the corporate income tax throughout the period 1999-2007 was 35% of profits. A lower tax rate of 30% was applied to firms under a revenue threshold that was modified over time: from €1.5 million in 1999 up to €10 million in 2010 (full details provided in Table 1). The eligibility cutoff for this tax break overlapped with the LTU threshold in 2004, but was different in every other year. The lower rate is applied only to the first €90,121 of profits (€120,202 since 2005) for eligible firms. Hence, there is notch for eligible firms with low profits, and a kink for those with high profits.

External Audit Threshold

Firms are required by law to have their annual accounts audited by an external private firm if they fulfill two of the following criteria for two consecutive years: (i) annual revenue above €4.75 million; (ii) total assets above €2.4 million;[23] and (iii) more than 50 employees on average during the year. These criteria also determine whether a firm can use the abbreviated form of the corporate income tax return, rather than the standard (long) version.

The external audit requirement complements tax enforcement policy because official tax audits typically use the private auditor's report as a source of information. Moreover, private auditors have an incentive to ensure truthful reporting because they face legal responsibility if any misreporting is found.[24] There is also a purely monetary cost associated with external audits: the fee charged by private audit firms is in the range €10,000 - €30,000 for firms with revenue close to €4.75 million, a small but non-negligible expenditure (0.2 to 0.6% of total revenue, but 4 to 12% of reported *profits* on average).

[23] The revenue limit was originally 790 million Pesetas (€4.748 million), and the assets limit was 395 million Pesetas (€2.374 million). These criteria were modified starting in 2008, raising the revenue threshold to €5.7 million and the assets threshold to €2.85 million.

[24] In private conversations, some auditors admit that they tolerate "small" amounts of misreporting, equivalent to about 2-3% of the firm's total operating revenue.

3.2 Data

All firms in Spain are required by law to submit a summarized version of their annual accounts to the Commercial Registry (*Registro Mercantil Central*). We use Amadeus, a European-level dataset that compiles all these financial statements, for the period 1999-2011.[25] The information available for each firm includes: name, location (5-digit postal code), sector of activity at the 4-digit level, 26 balance sheet items, 26 profit and loss account items, and 32 standard financial ratios.

The variable we use to determine whether firms are eligible to the LTU and other policies is *sales*, defined as the annual net revenue from sales. We also look at production inputs: *materials* includes the monetary cost of all raw materials and services purchased by the firm in the production process, and *costs of employees* measures the total wage bill of a firm, including social security contributions. The *number of employees* is reported at the end of the fiscal year (which coincides with the calendar year). This variable is missing for about 20% of the firms that report their total sales and material inputs.

The main advantage of this dataset is its longitudinal structure, which allows us to study the behavior of firms over time. Another important aspect is that we can analyze firm behavior both as taxpayers and as tax intermediaries along several dimensions, e.g. different tax bases, from a single data source. This is sometimes not possible with administrative tax returns, because these are usually anonymized and hence firms cannot be linked across different data sources. One limitation of the dataset is that many small firms do not submit their accounts to the Commercial Registry, because it is costly to them and the associated fines are small. Since the focus of this study is on firms with reported revenue around €6 million, we worry mostly about this group. Table 6 shows that more than 80% of firms with reported revenue between €3 and €9 million are included in Amadeus, which is reasonably close to the universe of all firms in this size range.

4 Empirical Strategy: Static Analysis

4.1 Operating Revenue Distribution

We begin by analyzing the distribution of firms' reported operating revenues. In the absence of any size-dependent regulation, our theoretical framework predicts a smoothly decreasing

[25] For the purposes of this paper, we accessed the online version of Amadeus in November 2011 for data on years 1999-2007 and April 2013 for the years 2008-2011. Since the dataset is continuously updated, the information currently available in the online version may have suffered some changes.

and convex density distribution. This is consistent with standard models of firm size determination (e.g., Lucas, 1978), and empirical regularities from comparable countries (e.g., Cabral and Mata, 2003). Any distortion in the reported revenue distribution at the thresholds described above indicates that there is a behavioral response to tax administration policies.

Using data from Amadeus, Figure 2 shows the distribution of reported revenues for Spanish firms in the range between €3 and €9 million for the period 1999-2007. We pool several annual cross-sections to increase the sample size and obtain smoother histograms, taking advantage of the fact that tax administration thresholds remained constant in nominal terms during this period. We observe two spikes in this distribution: a large one below the LTU threshold, and a much smaller one below the External Audit threshold. These behavioral responses indicate that firms are willing to incur a cost to report lower revenue in order to avoid entering the LTU census. The spike at the External Audit threshold is harder to interpret because the criteria to determine eligibility for the external audit involve other variables apart from reported revenue. For this reason, in the remainder of this paper we focus almost exclusively on the response to the LTU threshold.

There are two plausible concerns about the interpretation of the graphical evidence obtained by pooling multiple annual cross-sections. First, there may be heterogeneity in the bunching response across years that gets hidden in the aggregate picture. We show in the Appendix and Table 13 that the distribution of reported revenue is stable for each year throughout the period 1999-2007. A second concern is that there could be other size-dependent policies that simultaneously affect firm behavior. Apart from the small response to the External Audit threshold discussed above, we are only aware of another such policy during this period: the corporate income tax break for small firms described above. The annual revenue distributions plotted in Figure 12 show no discernible bunching at the tax break threshold in any year other than 2004, which is the only year when the two thresholds coincide. The lack of reaction to a 5 percentage-point reduction in the corporate income tax rate is striking in a context where firms are responding strongly to a discontinuity in tax enforcement intensity. This indicates that the existence of the LTU generates substantial incentives for firms to re-optimize.

4.2 Quantifying the Bunching Response

In order to quantify the size of the reported revenue response to the LTU cutoff, we use techniques from the bunching literature cited in the introduction. The key idea is to construct a counterfactual revenue distribution to estimate the excess bunching mass near the tax enforcement threshold. To do this, we fit a high-degree polynomial to the observed density, excluding an interval around the threshold where manipulation is most likely to occur. We call this interval the "excluded region" and we explain below how we determine its upper and lower bounds. As a first step, we divide the data in small bins of width w[26] and estimate the following polynomial regression:

$$F_j = \sum_{i=0}^{q} \beta_i \cdot (y_j)^i + \sum_{k=y_{lb}}^{y_{ub}} \gamma_k \cdot \mathbb{1}\,(y_j = k) + \eta_j \tag{8}$$

where F_j is the number of firms in bin j, q is the order of the polynomial, y_j is the revenue midpoint of bin j, the interval $[y_{lb}, y_{ub}]$ corresponds to the excluded region, and the γ_k's are intercept shifters for each of the bins in the excluded region.[27]

We estimate the counterfactual distribution by calculating predicted values with the estimated coefficients from regression (8), excluding the γ_k shifters to eliminate the perturbations around the threshold. Hence, the counterfactual density is given by:

$$\widehat{F_j} = \sum_{i=0}^{q} \widehat{\beta}_i \cdot (y_j)^i \tag{9}$$

Comparing this counterfactual density to the observed distribution allows us to estimate the excess bunching mass to the left of the threshold (B), and similarly the missing mass to the right of the threshold (H):[28]

$$\widehat{B} = \sum_{j=y_{lb}}^{y^{LTU}} \left|F_j - \widehat{F_j}\right| \qquad\qquad \widehat{H} = \sum_{j=y^{LTU}}^{y_{ub}} \left|F_j - \widehat{F_j}\right| \tag{10}$$

Determining the lower and upper bounds of the excluded region in a consistent way is critical for this estimation method to provide credible estimates. We follow the approach

[26] We use a bin width of €42,070, which allows us to precisely match the bin limits to each of the tax administration thresholds.

[27] In this particular application, we add to equation (8) dummy variables for the bins in the interval €4.5-€4.8 million, just below the External Audit threshold. This prevents the small spike in the density in that range from affecting the estimation of the counterfactual density around the LTU threshold. Adding these terms improves the accuracy of the counterfactual estimation around the LTU threshold as long as the bunching at the External Audit threshold is strictly local (i.e., firms bunching at the External Audit threshold would have had reported revenues just a little above it), which we believe is a reasonable assumption.

[28] We use absolute values to ensure that both estimates yield positive numbers. Otherwise, $\widehat{H}$ would be a negative number. Recall that $y^{LTU} =$ €6 million in our setting.

of Kleven and Waseem (2013), which is based on the principle that the area under the counterfactual density has to equal the area under the observed density. We start by setting an arbitrary lower bound, y_{lb}, and then run equation (8) multiple times. The idea is to eyeball the point where the distribution becomes distorted due to the bunching response, since revenue manipulation is usually imprecise and not all bunching firms manage to locate exactly at the threshold. Regarding the upper bound, in the first iteration we set $y_{ub} \approx y^{LTU}$, which tends to yield large estimates of $\widehat{B}$ and small estimates of $\widehat{H}$. The estimation routine is programmed to increase the value of y_{ub} by a length w and run equation (8) again as long as $\widehat{B} > \widehat{H}$. The process continues until it reaches a value of y_{ub} such that $\widehat{B} = \widehat{H}$.[29]

The results obtained allow us to estimate the bunching parameter b defined in equation (7), which equals the ratio of excess bunching mass over the average height of the counterfactual density in the interval $\left(y_{lb}, y^{LTU}\right]$. The actual estimator formula is given by:

$$\widehat{b}_{NF} = \frac{\widehat{B}}{\left[\frac{1}{1+(y^{LTU}-y_{lb})/w}\right] \sum_{j=y_{lb}}^{y^{LTU}} \widehat{\beta}_i \cdot (y_j)^i}, \tag{11}$$

where the term $\left[1 + \left(y^{LTU} - y_{lb}\right)/w\right]$ is the number of excluded bins below the threshold. We use the subscript NF to indicate that it is defined under the assumption of no optimization frictions. This assumption implies that every firm has the ability to modify its reported revenue as it wants (through real or evasion responses) in order to bunch below the threshold. The assumption is very restrictive in this setting, since we can see in Figure 2 that many firms report revenues just above the LTU threshold. We discuss a correction to this estimator that takes optimization frictions into account in the next subsection.

Since this estimation procedure is applied to the universe of Spanish firms rather than a random sample, there is no sampling error and therefore we cannot construct the usual confidence intervals. To test whether the point estimates are statistically significant, we sample the residuals from regression (8) a large number of times (with replacement) to obtain bootstrapped standard errors.[30]

We obtain a point estimate of $\widehat{b}_{NF} = 0.086$ ($s.e.$ 0.005) for the period 1999-2007 and $\widehat{b}_{NF} = 0.026$ ($s.e.$ 0.004) for the period 2008-2011. Both are precisely estimated and statis-

[29] Recall that w is the width of the bins used to build the counterfactual. The fact that there is a finite number of bins means that, in practice, we need to impose the weaker condition that the ratio is "close" to one: $\left(\hat{H}/\hat{B}\right) > 0.95$.

[30] We thank Michael Best for sharing his Stata code to perform this bootstrapping routine. In all the results shown below, we perform 200 iterations to obtain the standard errors. Using a larger number does not affect our results.

tically different from zero at the 1% level. To interpret the estimator $\widehat{b}_{NF}$, we make two key assumptions. First, we assume that firms face no optimization frictions, as explained above. Second, we assume that the smoothly decreasing counterfactual density defined by (9) is a good approximation of the theoretical revenue distribution in the absence of the LTU threshold. Under these assumptions, the results for 1999-2007 mean that the marginal buncher reports revenue €86,000 lower, or 1.4% of their total revenue, than it would have if the LTU threshold did not exist (€26,000, equivalent to 0.4% of total revenue, for 2008-2011).

Most papers in the bunching literature (e.g., Saez, 2010; Chetty et al., 2011) use b as the numerator of the elasticity of taxable income, the structural parameter of interest in their settings. The denominator in that elasticity is the proportional change in the net-of-tax rate.[31] In our setting, the policy that changes at the threshold is the probability of detection δ, which is very difficult to measure because enforcement strategies include many elements (audit probabilities, ability to cross-check transactions, etc.) that are themselves hard to quantify. Therefore, we do not attempt to estimate the elasticity of reported income with respect to tax enforcement, which would be the structural parameter of interest. This does not mean that our results cannot be generalized to other contexts. Dozens of countries around the world have established LTUs within their tax agencies and, although the designs vary widely, many of them also use revenue thresholds to determine eligibility OECD (2011). Our results are therefore indicative of the potential effects of setting LTU eligibility thresholds based only on reported revenue.

To address the concern that the arbitrary selection of y_{lb} could bias the estimation, we perform a sensitivity analysis. We pick different values for the lower bound of the excluded region around our preferred value of €5.5 million, such that $y_{lb} = \{5.3, ..., 5.7\}$. Table 3 reports the results for the pooled 1999-2007 data. The upper bound y_{ub} is quite stable between €6.5 and €6.6 million. Similarly, point estimates for $\widehat{b}_{NF}$ are always between 0.081 and 0.086. One of the reasons why these estimates are so robust is that the revenue distribution for the period 1999-2007 is very smooth except in the interval around the threshold, where bunching is substantial. When applying the same method to distributions with less bunching or more noise, the estimates tend to be more sensitive to the choice of y_{lb}. The same is true of regression analysis when the variance of the dependent variable is very high compared to that of the explanatory variables and the researcher specifies different functional forms.

[31] The net-of-tax rate is defined as $1 - t$, where t is the tax rate.

4.2.1 Optimization Frictions

Contrary to prediction of the stylized model without frictions, we do not observe a hole in the distribution just above the LTU threshold – just a small dip. This suggests that some firms are not able to adjust their reported revenue as easily as others, and end up reporting revenues just above the cutoff. Thus, the monetary interpretation of estimates of b_{NF} may not a precise measure of firms' structural response to a change in tax enforcement, because it ignores the influence of optimization frictions on the behavior of some firms.

Optimization frictions have been a widely discussed issue in the bunching literature, sometimes because the cost of not re-optimizing is low in many contexts. This is particularly relevant at kink points, where the marginal tax rate jumps discontinuously but the average tax rate varies smoothly. For example, Chetty (2012) shows that an adjustment cost equivalent to 1% of total expenditure makes a high intensive-margin elasticity compatible with a zero bunching response. The incentives to bunch are considerably stronger in the case of notches, because the associated cost of inaction grows at a first-order rate with the size of the policy change (Slemrod, 2010; Chetty, 2012).

Even though businesses have more control over their reported income than wage earners (whose income is third-party reported), there are several reasons why firms might not respond to the existence of the LTU. First, about half of the firms locating just above the cutoff in any given year had previous exposure to the LTU. That is, their revenue had already been above €6 million for at least one year before the moment in which we observe them. Second, some firms may not be planning to misreport their activities regardless of the enforcement regime. This could be due to preferences of the manager against tax evasion or perhaps due to inability to evade given some sector characteristics (e.g., government contracts). For these firms, the only consequence of being in the LTU is facing additional compliance costs. Third, firms might be unable to control their revenue with precision due to adjustment costs or unexpected shocks. Fourth, as mentioned in the previous section, exporters are always included in the LTU regardless of their revenue, so they do not have incentives to manipulate their revenues to avoid the additional tax enforcement.

We illustrate the importance of the first reason with some evidence for growing and shrinking firms. Recall that firms are added to the LTU census the year after their revenues rise above €6 million, and they are taken out one year after their revenues drop below the cutoff. Despite this formal symmetry, entering the LTU in practice forces some businesses to make important administrative changes to adapt to the higher enforcement regime. For example, they would have to give up having two sets of accounting books. Once the firm puts an end to the parallel accounting system, it is hard to set it up again after dropping out of the LTU census. Moreover, in small regions there is only a few hundred large firms, which are well known by the local LTU staff. Anecdotally, officers from the tax agency report that marginal firms in small regions often move their headquarters to a large city (e.g., Madrid, the capital) to blend into a larger group of firms and lower their expected probability of audit.

To test whether entering the LTU is seen by firms as a fixed cost, we compare the behavior of firms whose reported revenue is growing to those that are shrinking. Specifically, a growing firm is defined as having higher revenues in year t than in $t-1$ (vice versa for shrinking). Figure 6 shows the striking differences in the revenue distributions for these two groups for the full period 1999-2011. Growing firms bunch very strongly at the LTU threshold, but barely react to the External Audit threshold. In contrast, shrinking firms do the exact opposite: they bunch in response to the External Audit requirement, but their response to the LTU cutoff is minimal.[32] We conclude that some growing firms avoid the LTU because they anticipate it will involve paying a one-time adjustment cost and it will reduce their ability to evade taxes in future years. In contrast, shrinking firms with previous LTU exposure have less to gain from bunching just below the threshold because they have already incurred the fixed cost.

Rather than introducing each source of rigidity explicitly into the model, we assess their combined impact to an upper bound of the structural response.[33] We define α as the proportion of firms locating in the interval $\left(y^{LTU}, y_{ub}\right]$, compared to the counterfactual density. This includes all firms that do not bunch even though there are firms similar to them (according to our counterfactual) that do bunch. We use this measure to re-weight the estimates of $\widehat{b}_{NF}$, and use the subscript F to indicate that the new estimator accounts for optimization frictions. Thus, $\widehat{b}_F$ can be thought of as treatment-on-the-treated estimator for firms with low adjustment costs:

$$\widehat{b}_F = \frac{\widehat{b}_{NF}}{1-\alpha} \tag{12}$$

[32] In a more disaggregated analysis, we observe that the only subset of shrinking firms that features bunching at the LTU threshold is composed of firms with revenue falling between 0% and -3%. However, firms with a revenue decrease of -3% or beyond show no bunching response. There is always some bunching at the External Audit threshold for these two groups.

[33] Kleven and Waseem (2013) propose a similar method to account for optimization frictions, although in their case there is a strictly dominated region in which no taxpayer should locate under any preferences, because their take-home pay falls as income rises. In our setting, there is no strictly dominated region because there may be heterogeneity in the optimization frictions faced by firms.

We interpret estimates of $\widehat{b}_F$ as an upper bound of the firms' response to a change in tax enforcement, since $\widehat{b}_F \geq \widehat{b}_{NF}$ by definition (notice that $\alpha \geq 0$). We calculate standard errors for this estimator with the same bootstrapping procedure used above.

The estimate taking frictions into account is $\widehat{b}_F = 0.465$ ($s.e.$ 0.052) for the period 1990-2007 and $\widehat{b}_F = 0.384$ ($s.e.$ 0.042) for 2008-2011. To provide a sense of the magnitude of this response, consider that the average profit margin of firms around the LTU threshold is 4.4% of revenue, approximately €290,000. If the entire response is due to revenue underreporting, then the marginal buncher's would wipe out its taxable profits completely and evade its entire tax liability. However, caution is warranted because we do not know to what extent the response is pure evasion or there is also a real response.

4.3 Heterogeneous Responses

We have shown that the annual revenue distributions are stable within the two broad periods of economic boom (1999-2007) and recession (2008-2011). However, there could be a great deal of heterogeneity across multiple dimensions such as: number of employees, fixed assets, organizational form, sector of activity, and region where the headquarters are located. The main results of these exercises are reported in Table 4 and Figures 17-5.

Heterogeneous Responses across other Dimensions of Firm Size. Conditional on being the neighborhood of the LTU threshold, firms with more employees and/or assets tend to have a more complex structure, so they need to have sophisticated accounting systems in place that make misreporting more costly and risky. Holding everything else constant, we expect to see the strongest bunching response among smaller firms along these dimensions. Figures 17 and 18 plot the revenue distributions for groups of firms of different sizes in terms of employees and fixed assets. The density distribution is strongly right-skewed for the smallest firms, while it is almost flat for the largest ones. Bunching at the LTU threshold is strongest among firms below the 50-employee and €2.4-million-in-assets marks.[34] The "no frictions" bunching estimates are in the same order of magnitude for the very small (less than 40 employees) and large firms (more than 50), with $\widehat{b}_{NF} \approx 0.08$. Similar results are found for firms with less than €5 million in assets, but the bunching estimates are much smaller and only marginally significant for firms with more assets.

[34] Notice that these are two of the three eligibility criteria in the External Audit threshold. There are also a number of labor regulations that apply only to firms with more than 50 employees, for example the obligation to have a Workers' Council where unions are represented and acquire some decision power within the company.

Table 4 also reports bunching estimates for firms with different legal forms, *"Sociedad Limitada (SL)"* (comparable to Limited Liability Company in the US) and *"Sociedad Anónima (SA)"* (comparable to a Corporation). The capital requirements to set up a SL are smaller than for a SA, but the latter is the natural legal form for publicly traded companies. SL's are more numerous and smaller on average, but we do not find significantly different bunching responses, as can be seen in Figure 19. This can be explained by the fact that both legal forms are treated equally in terms of taxation.

Regional Variation. Given that the LTU is organized in regional offices, there might be variation in the enforcement intensity change experienced when crossing the threshold in each region. Figure 20 shows a map with the 17 Autonomous Regions in Spain. We use a color scale to show the different bunching intensity observed in the revenue distribution in each region. Lighter (yellow) tones apply to low bunching regions, while darker tones (red) denote high-bunching regions. The lowest bunching is observed in Navarra and País Vasco, the two regions in the North-Center where the Large Taxpayers' Unit (LTU) only applies to firms those that operate extensively in the rest of Spain. There is relatively (but statistically significant) low bunching in the Northern and Eastern regions of Cataluña, Aragón, Valencia and Baleares. Meanwhile, bunching is relatively high in the South, Center and North-West. The top bunchers are Extremadura, a relatively poor region in the Center-West, and Cantabria, a middle-income small region in the North. One potential story is that the prevalence of tax evasion is higher in the regions with larger bunching responses. Alternatively, it could be that firms have stronger incentives to bunch in regions where the LTU office is more competent. Since we do not have reliable measures of the quality of tax enforcement in each regional LTU office, it is difficult to provide an clear interpretation of this regional heterogeneity.

Heterogeneity across Sectors of Activity. Firms in different sectors of the economy face different constraints on misreporting, depending on their technology, e.g. whether they are labor intensive or not. A restaurant with €6 million in revenue is typically a medium-large company with dozens of employees, and most likely with more than one location. In contrast, a merchant wholesaler that sells electronic products typically reaches that revenue level about 15 employees. To explore how companies operating in such different markets respond to the same nominal revenue threshold, we define 12 different sectors of economic activity (details on how the sectors are defined can be found in the Appendix). The intuition is that selling to final consumers makes underreporting much easier because the VAT self-enforcing mechanism breaks down at that stage. For the second element, the idea is that it is easier to underreport if the number of employees is small.

Figure 5 shows a scatterplot of bunching responses by sector (measured by $\widehat{b}_{NF}$, vertical axis) and final consumption as a share of total sales (horizontal axis). The relationship is downward slope, suggesting that the incentives to remain under the LTU threshold are stronger for firms in sector where a low percentage of sales is made to final conumers. One possible theory to explain this trend is that when retailers underreport, it is extremely hard even for the LTU to detect it because it hard to trace small cash transactions. Moreover, the median number of employees of retail firms around the revenue cutoff is just 21. Between these two extremes we observe mostly the industrial sectors: manufacturing, wholesale and construction. Focusing only on these sectors, it seems that the relationship between scope of evasion and bunching intensity could be positive, but the small number of data points limits our ability to draw a strong conclusion.

5 Empirical Strategy: Dynamic Analysis

The analysis from the previous section imposes a static perspective by pooling observations from different years. This means that many firms appear in the data in multiple years, but the graphical analysis does not control for potential autocorrelation. To improve our interpretation of the bunching response, we are interested in understanding the dynamic behavior of firms. In particular, we are concerned that persistence in behavior could bias our bunching estimates. We present below some descriptives of firms' growth patterns and analyze the extent of bunching persistence.

Figure 7 shows the proportion of firms whose revenue grows between years t and $t + 1$, plotted against reported revenue in year t. For the boom period 1999-2007 (top), about 61% of firms on average increase their revenue the following period. This proportion is stable in the range €2-€10 million with a slight upward trend, except for a short interval below the LTU threshold where it drops significantly to about 55%. The drop is less pronounced for the recession period (2008-2011), where the proportion of firms growing is substantially lower at around 37% (bottom panel). These patterns broadly corroborate the intuition from the static estimates: as small firms grow and approach the threshold, a subset of them slows down their growth to avoid crossing it. Notice that it is enough with a small subset of firms reverting their trend to generate a substantial amount of bunching.

Another interesting variable to look at is median revenue growth in the following year, that is $\Sigma_i median\left(ln\left(y_{i,t+1}\right) - ln\left(y_{i,t}\right)\right)$. We consider median instead of average growth because the distribution of growth rates has very long tails of extreme values and hence bin averages are erratic.[35] The top panel of Figure 7 shows median revenue growth by current revenue bins for the boom period. Growth rates are similar for firms with current revenue between €2-€5 and €6-€10 million. However, there is a sharp decline in median growth rates as firms approach the threshold from below, i.e., between €5-€6 million. This is another indication that a group of firms artificially reduces its growth as they approach the threshold. The pattern is smoother in the recession period (bottom panel), with much lower median growth rates, as shown in the bottom panel of Figure 7.

Finally, we consider persistent bunching behavior, which could bias our static estimates if our bunching estimator is only capturing the behavior of a few firms that remain just below the threshold for many years. It is important to keep in mind that the LTU notch was fixed in nominal terms throughout the 1999-2011 period, but inflation averaged 3% per year. Thus, the notch moved down about 27% between 1999 and 2007, and 43% if we consider the full period up to 2011.[36] To obtain a measure of bunching persistence, we follow the approach proposed by Marx (2012). The idea is to estimate whether firms are more likely to stay in the bunching region than in any other part of the revenue distribution. In order to precisely define the bunching region, we divide reported revenues in equally sized bins of €250,000, such that the bunching bin includes reported revenues between €5.75 and €6 million.[37] We then compare the fraction of firms that remain in the bunching bin after h years to the fraction that remain in other revenue bins, where $h = \{1, 2, ..., 10\}$. In the data, the probability of staying in a given revenue bin decreases with revenue for all values of h, because the equal-sized bins are proportionally smaller as we move to higher revenue levels. We estimate the following regression model:

$$Prob\left[bin\left(y_{it}\right) = bin\left(y_{i,t+h}\right)\right] = \alpha + \beta BunchBin_{it} + y_{it} + y_{it}^2 + \lambda_t + \varepsilon_{it} \tag{13}$$

where y_{it} is reported revenue by firm i in year t, the left-hand side variable is the fraction of firms that report revenues in the same bin in years t and $t+h$, $BunchBin_{it}$ takes value one if $y_{it} \in (5.75, 6]$, and λ_t denotes a vector of year dummies. We add a quadratic polynomial in current reported revenue as a way to control for the counterfactual probability that firms

[35] In particular, there is a considerable number of firms whose revenues drop from a few million euros to basically zero the following year. The large negative growth rates registered by these firms bias average growth rates down, resulting in negative numbers even during the boom years.

[36] We obtain these numbers simply calculating $(1.03)^8 = 1.27$ and $(1.03)^{12} = 1.43$.

[37] The results are qualitatively similar for other bin widths, such as €100,000 or €500,000.

remain in a given revenue bin. Instead of using revenue levels, we use the distance to the notch so that the constant term α can be interpreted as the fraction of firms near the notch expected to remain at their current revenue level h years from now.

Figure 8 shows the results graphically. The top-left diagram shows the probability that firms remain in the same revenue bin after one year. This probability decreases smoothly from about 24% in the range $y_{it} \in$ (€2, €2.25 million) to 6% in the range $y_{it} \in$ (€9.75, €10 million). However, there is a clear deviation from the trend at the bunching bin, where the proportion of firms that stay is 14.8%, compared to the 8.4% predicted by the counterfactual. This means that a firm in the bunching bin is 75 *percent* (6.3 percentage points) more likely to remain in the same revenue bin one year later. The regression results for all values of h are summarized in Table 5. The coefficient on the $BunchBin$ dummy is significant at the 5% level for all lags up to six years, although the economic significance is much stronger for the short lags (up to three years).

6 Empirical Analysis: Reported Input Costs

In the appendix, we derive several extensions to the model from Section 4 to allow firms to choose their input use and also to misreport their input costs. Here, we summarize briefly the theoretical predictions for the model with two inputs (materials and labor) and two taxes (corporate income tax and VAT) and present some empirical evidence. When bunching is due to a real production response, this model predicts an upward jump in reported input costs at the threshold (that is, firms just below the threshold report less labor and material inputs than firms just above). This follows from the fact that bunching firms have higher productivity and the production function is concave, so bunching firms need fewer inputs to produce the same output as other firms. When bunching is due to tax evasion, the model predicts a downward jump in reported material inputs at the LTU threshold. This is due to bunching firms producing more output (and hence using more inputs) than they report. Since we expect bunching firms to also overreport their material inputs, the jump will be even more pronounced. In contrast, the model predicts a downward jump in reported labor costs, for exactly the opposite reasons.

These predictions can be tested with simple graphical evidence showing how the reported input ratios behave around the LTU threshold. We use these graphs to rule out some of the stories consistent with the models, rather than to identify causal effects. Recall that firms are included in the LTU census the year after they cross the threshold. Therefore, we could interpret that the degree of tax enforcement (low or high) in a given year depends on reported revenue the previous year. This interpretation implicitly assumes that firms do not know whether they will finish the year above or below €6 million in reported revenue. However, it can also be argued that firms are likely to anticipate what type of enforcement they will face the following year, and hence they will adjust their reporting behavior accordingly. Since there are good arguments in favor of both interpretations, we show all our results considering the outcome variables (reported input costs) in year $t+1$ and also in year t, always against revenues in year t.

6.1 Graphical Evidence

The left panels of Figure **??** plot the ratio of reported input costs over reported revenue on the vertical axis and reported revenue in the horizontal axis, both measured in year t. The right panels plot the same variables, but in this case the ratio of inputs over revenue measured in period $t+1$. The solid lines show a quadratic fit of bin averages with 95% confidence intervals, while the dots indicate median values for each bin. All bins are €200,000 wide. The top panels include data from the boom period (1999-2007) and the bottom panels for the recession period (2008-2011). There is no adjustment for inflation in any of the graphs because the outcome variable is a ratio of two nominal amounts.[38] The ratio of inputs over revenue is remarkably stable for different levels of revenue at approximately 94% for all variable definitions and periods. Both medians and averages show a small downward jump at the LTU threshold, but the difference is statistically insignificant so the evidence is inconclusive with respect to the models' predictions.

The same four plots are shown in Figure 9 using the ratio of material input costs over revenue as the outcome variable. The ratio slopes up in a concave shape, indicating that firms with higher revenues use an increasingly higher proportion of material inputs. In the boom period, the material over revenue ratio jumps sharply downwards by about 1 percentage point (the median value for the ratio is around 70% in the boom period and 66% in the recession period). This is true both for bin averages and medians, and the distance is statistically significant. The pattern is similar for the recession period, but the jump is smaller and not significant. These patterns are compatible with an evasion response where firms either underreport their revenue or overreport their materials. In contrast, they are incompatible with a fully real response, because in that case we should observe an upward jump at the threshold.

[38] We assume implicitly that the inflation on the output good is the same as for inputs.

Figure 10 shows the same evidence for the ratio of labor inputs over revenue. The patterns observed are approximately the inverse of those for materials: there is an upward jump in the ratio of labor costs over revenue at the threshold, which is more pronounced during the boom period than the recession period. The size of the jump is approximately 1 percentage point, but in this case it is more relevant because median labor costs are about 12% of revenue. The upward jump is compatible both with a real response (highly productive bunching firms need less labor to produce the same output) and with an evasion response in which labor costs are underreported.

There are two broad interpretations for these patterns of materials and labor input costs around the LTU threshold. First, it could be that bunching firms respond to the differential tax incentives by overreporting material inputs and underreporting labor expenses. We call this the "input-misreporting hypothesis". It is consistent with the theory for the two jumps to cancel each other out and thus not lead to any discontinuity in total reported inputs. A second interpretation is that labor-intensive firms are less likely to bunch below the LTU threshold, which mechanically yields lower average labor inputs in the bunching interval. We call this the "composition-effect hypothesis". Under this hypothesis, the discontinuities at the threshold would be explained by differential sorting across sectors. The two interpretations are observationally equivalent, so we need additional tests to determine which hypothesis is more plausible.

Figures 11 and 11 provide a more disaggregated picture of labor input costs. Figure 11 plots our measure of average wages, which features an upward jump at the threshold. The jump is more pronounced and statistically significant for the boom period, as was the case for labor inputs. It is harder to visualize a discontinuity in the average or median number of employees at the threshold, as can be seen in Figure 11. The fact that the drop in labor costs is mostly due to lower average wages (rather than fewer employees) seems easier to square with the evasion hypothesis. To associate this with the composition-effect hypothesis, one would have to explain why less labor-intensive firms also pay lower wages on average. In any case, the evidence presented in this subsection is only suggestive and is not enough to assert with full certainty that the evasion hypothesis is correct.

7 Efficiency Costs of Tax Enforcement

The empirical results obtained in the previous sections suggest that firms respond to the tax enforcement threshold mostly by underreporting their operating revenue, without reducing actual production in a significant way. Drawing from the literature on the deadweight loss of taxation in the presence of evasion and avoidance, we provide an upper bound for the efficiency costs of tax enforcement. In the final subsection, we perform a rough calculation of the losses in tax revenue due to evasion in the low enforcement regime.

As noted above, the introduction of a proportional tax on profits did not generate an inefficiency in this framework, but the distortions created by tax enforcement elicit behavioral responses from firms that could lead to efficiency costs. To set up a social welfare function, we make the simplifying assumption that each firm is owned by one individual, whose total income is the after-tax profit of the firm. That way, we can aggregate these individuals' welfare to the tax revenue raised by the government and make meaningful comparisons. In this theoretical framework, an increase in tax enforcement (summarized by the probability of detection, δ) is equivalent to an increase in the expected tax rate. Therefore, we can evaluate how expected welfare changes in response to an increase in tax enforcement in the same way that the literature on the deadweight loss of taxation evaluates the welfare implications of tax changes.

We define expected welfare as the sum of expected profits and expected tax revenue:

$$\mathbb{E}W(\delta) = \{(1-t)\Pi + tu[1-\delta(1+\theta)]\} + t[\Pi - u[1-\delta(1+\theta)]] \tag{14}$$

By envelope theorem, we can ignore behavioral responses in the term in curly brackets, because firms are already choosing m and u to maximize expected profits. Hence, an increase in tax enforcement leads to the following change in expected welfare:

$$\begin{aligned}
\frac{d}{d\delta}\mathbb{E}W &= t\left[-u(1+\theta) + \frac{d\Pi}{d\delta} - \frac{du}{d\delta}[1-\delta(1+\theta)] + u(1+\theta)\right] \\
&= t\frac{d\Pi}{d\delta} - t\left[\frac{du}{d\delta}[1-\delta(1+\theta)]\right] && (15) \\
&= t\frac{d\overline{\Pi}}{d\delta} + t\frac{du}{d\delta}[\delta(1+\theta)] && (16)
\end{aligned}$$

We know from comparative statics that $\frac{du}{d\delta}|_{u=u^*} < 0$, so the second term in (16) is negative. This implies that the change in expected welfare due to an increase in enforcement is neither proportional to the elasticity of reported taxable profits $\left(\overline{\Pi}\right)$ nor to the elasticity of true profits (Π), but to an intermediate amount. Formally,

$$t\frac{d\Pi}{d\delta} \leq \frac{d}{d\delta}\mathbb{E}W \leq t\frac{d\overline{\Pi}}{d\delta}. \tag{17}$$

Hence, the efficiency cost of tax enforcement cannot be calculated based solely on the effects on reported profits. The response of true profits has to be taken into account as well. We return to this discussion in Section 6, after presenting our empirical results.

A crucial question for the design of tax administration policies is whether there are large efficiency costs from tax enforcement. The previous subsection laid the ground for this estimation by deriving expressions for the change in expected welfare associated to an increase in tax enforcement. In our theoretical framework, an increase in tax enforcement is equivalent to an increase in the expected tax rate.

In two influential papers, Feldstein (1995, 1999) argued that the elasticity of taxable income with respect to tax rate changes is a sufficient statistic to estimate the excess burden of taxation. This result is useful because it accounts for all the key behavioral responses to taxation (labor supply, avoidance, and evasion), and also because taxable income data is widely available. The key assumption driving Feldstein's result is that tax evaders equate the marginal private cost of evasion (or avoidance) to the marginal cost of reducing true income (by producing less), so that the specific reason why they report lower income does not matter for efficiency.

Chetty (2009) points out that Feldstein's result implicitly assumes that the marginal social cost of evasion and avoidance differs from the tax rate (the marginal benefit). Chetty considers two types of sheltering costs (where sheltering includes both evasion and avoidance). First, there are "resource costs" that make production less efficient when there is evasion. For example, the need to have accountants keeping two different books, or the lost profits for operating in cash. If evaders only incur a resource cost, then Feldstein's result holds. A conceptually different cost is what Chetty calls "transfer costs", for example a monetary penalty to punish evasion behavior. A penalty has a private cost to the evader, but no social cost because the resources are transferred to the government another agent (assuming risk neutrality, as is standard when modeling firm behavior). Chetty shows that the excess burden of taxation in the presence of such transfer costs is directly proportional to the elasticity of total earned income (as opposed to taxable income).

The theoretical framework presented in Section 2 is slightly more complicated than Chetty's because the probability of detection depends on firm size, besides the amount evaded. In spite of this difference, equations (15) and (16) deliver a qualitatively similar result: when firms face only transfer costs of evading, the deadweight loss generated by an increase in tax enforcement is less than proportional to the effect on reported profits. The lack of a real production response to the existence of the LTU implies that the efficiency cost of increasing tax enforcement is not high in this context. The effect of this enforcement policy is to redistribute resources from tax-compliant firms (or other taxpayers) to tax evaders.

One aspect that we have not addressed so far is the administrative cost of higher enforcement. This could be easily introduced in the model with the function $q(\delta)$, which is increasing in δ. The modified equations (15) and (16) would be:

$$\frac{d}{d\delta}\mathbb{E}W = t\frac{d\Pi}{d\delta} - t\left[\frac{du}{d\delta}\left[1-\delta\left(1+\theta\right)\right]\right] - \frac{dq}{d\delta} \tag{18}$$

$$= t\frac{d\overline{\Pi}}{d\delta} + t\frac{du}{d\delta}\left[\delta\left(1+\theta\right)\right] - \frac{dq}{d\delta} \tag{19}$$

It is challenging to obtain measures of the marginal increase in administrative costs associated to an increase in tax enforcement. We only have access to the total cost of the tax agency in Spain, which was €1.33 billion in the year 2007, when it raised €250 billion in tax revenue (€188 collected per €1 spent or, equivalently, 0.5 cents of a euro per each euro collected).[39] The marginal return to spending an additional euro on enforcement is likely to be below this average return, but also well above an additional euro in tax revenue. To guide a cost-benefit calculation to determine what is the socially optimal enforcement intensity, we would need to consider the tax revenue lost due to low enforcement.

8 Concluding Remarks

This paper examines the behavioral responses of firms to increases in tax enforcement. The empirical analysis relies on a theoretical framework that considers firms' incentives to evade taxes. The predictions of this model are tested using the discontinuity in tax enforcement intensity created by a Large Taxpayers Unit in Spain.

The empirical results show a strong reaction of firms to avoid larger tax enforcement that creates a significant distortion in the firm size distribution in Spain. This persistent pattern over time is created by groups of firms that as they approach to the threshold reduce their reported revenue increasing hidden activity. The documented heterogeneous reaction among firms illustrates the relevance of several features, such as the position of the firm in the productive firm or the number of employees, to determine firms' ability to evade taxes. The

[39] The Internal Revenue Service (IRS) of the United States, considered one of the most efficient tax agencies in the world, collected \$2.4 trillion in 2007, with an administrative cost of \$10.7 billion. Hence, the IRS collects \$224 per \$1 spent, higher than the Spanish tax agency, but in the same order of magnitude. (Source: www.irs.gov).

patterns of the data also point that firms staying in the low enforcement regime have larger scope to evade taxes by overreporting their material input costs and also colluding with their employees to pay part of their wages “in black” (potentially using the underreported revenue).

The results of this paper reveal that third-party reporting by firms and the presence of paper trail in the VAT are not enough to guarantee complete tax enforcement. Tax agencies must devote resources to closely monitor firms undertaking cross-checking and audits that benefit from transactions information provided by taxpayers. Overall, the conclusions of the paper imply that theoretical and empirical research on tax evasion must put more emphasis on examining the behavior firms, the critical fiscal agent in advanced economies.

References

AEAT, "Memoria de Administración Tributaria," Technical Report, Agencia Estatal de Administración Tributaria (Spanish Tax Agency) Several years.

Allingham, Michael and Agnar Sandmo, "Income tax evasion: A theoretical analysis," *Journal of Public Economics*, 1972, *1*, 323–338.

Bastani, Spencer and Hakan Selin, "Bunching and Non-Bunching at Kink Points of the Swedish Tax Schedule," June 2012. CESifo Working Paper No. 3865.

Bayer, Ralph and Frank Cowell, "Tax compliance and firms' strategic interdependence," *Journal of Public Economics*, 2009, *93*, 1131–1143.

Cabral, Luís and José Mata, "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review*, September 2003, *93* (4), 1075–1090.

Chetty, Raj, "Is the Taxable Income Elasticity a Sufficient Statistic to Calculate Deadweight Loss? The Implications of Evasion and Avoidance," *American Economic Journal: Economic Policy*, 2009, *1* (2), 31–52.

_ , "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, May 2012, *80* (3), 969–1018.

_ , John Friedman, and Emmanuel Saez, "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings," July 2012. NBER Working Paper 18232.

_ , _ , Tore Olsen, and Luigi Pistaferri, "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *Quarterly Journal of Economics*, 2011, *126*, 749–804.

Christensen, Kevin, Robert Cline, and Tom Neubig, "Total Corporate Taxation: "Hidden", Above-the-Line, Non-Income Taxes," *National Tax Journal*, 2001, *LIV* (3), 495–506.

Devereux, Michael, Li Liu, and Simon Loretz, "The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records," January 2013. Oxford University Centre for Business Taxation Working Paper 12/23.

Dharmapala, Dhammika, Joel Slemrod, and John Douglas Wilson, "Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs," *Journal of Public Economics*, October 2011, *95* (9-10), 1036–1047.

Feldstein, Martin, "The Effect of Marginal Tax Rates on Taxable Income: A Panely Study of the 1986 Tax Reform Act," *Journal of Political Economy*, June 1995, *103* (3), 551–572.

_ , "Tax Avoidance and The Deadweight Loss of Income Taxation," *Review of Economics and Statistics*, November 1999, *81* (4), 674–680.

IEF, "Evolución del Sistema Fiscal Español: 1978-2010," 2011. Instituto de Estudios Fiscales.

IMF, "Improving Large Taxpayers' Compliance: A Review of Country Experience," Technical Report, International Monetary Fund, Washington DC 2002. Team led by Katherine Baer.

Keen, Michael and Stephen Smith, "VAT Fraud and Evasion: What do we know, and what can be done?," Working Paper, IMF Fiscal Affairs Department February 2007.

Kleven, Henrik and Mazhar Waseem, "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *forthcoming at the Quarterly Journal of Economics*, 2013.

_ **, Klaus Kreiner, and Emmanuel Saez**, "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries," August 2009. NBER Working Paper 15218.

Kumler, Todd, Eric Verhoogen, and Judith Frias, "Enlisting Workers in Monitoring Firms: Payroll Tax Compliance in Mexico," June 2012.

Lucas, Robert, "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, Autumn 1978, *9* (2), 508–523.

Marx, Benjamin, "Regulatory Hurdles and Growth of Charitable Contributions: Evidence from a Dynamic Bunching Design," November 2012. Columbia University.

OECD, "Tax Administration in OECD and Selected non-OECD Countries: Comparative Information Series (2010)," Technical Report, Organisation for Economic Co-operation and Development March 2011.

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

Paula, Áureo De and José Scheinkman, "Value-Added Taxes, Chain Effects, and Informality," *American Economic Journal: Macroeconomics*, October 2010, *2*, 195–221.

Pomeranz, Dina, "No Taxation Without Information: Deterrence and Self-Enforcement in the Value-Added Tax," April 2013. Harvard Business School Working Paper 13-057.

Reinganum, Jennifer and Louis Wilde, "Income Tax Compliance in a Principal-Agent Framework," *Journal of Public Economics*, 1985, *26*, 1–18.

Restuccia, Diego and Richard Rogerson, "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 2008, *11*, 707–720.

Saez, Emmanuel, "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, August 2010, *2* (3), 180–212.

Schneider, Friedrich, Andreas Buehn, and Claudio Montenegro, "Shadow Economies all over the World: New Estimates for 162 Countries from 1999 to 2007," Working Paper 322, Universidad de Chile 2010.

Slemrod, Joel, "The economics of corporate tax selfishness," *National Tax Journal*, December 2004, *LVII* (4), 877–900.

_ , "Does it Matter Who Writes the Check to the Government? The Economics of Tax Remittance," *National Tax Journal*, June 2008, *LXI* (2), 251–275.

_ , "Buenas Notches: Lines and Notches in Tax System Design," 2010. Working Paper, University of Michigan.

Tables

Table 1: Revenue Threshold: Corporate Income Tax Benefit for Small Firms

<table>
<tr><th>Year</th><th>Threshold</th><th>Standard tax rate</th><th>Special tax rate</th><th>Applicable range</th></tr>
<tr><td>1999</td><td rowspan="2">€1.5 million</td><td rowspan="8">35%</td><td rowspan="8">30%</td><td rowspan="6">Up to €90,151 of taxable profits</td></tr>
<tr><td>2000</td></tr>
<tr><td>2001</td><td>€3 million</td></tr>
<tr><td>2002</td><td rowspan="2">€5 million</td></tr>
<tr><td>2003</td></tr>
<tr><td>2004</td><td>€6 million</td></tr>
<tr><td>2005</td><td rowspan="5">€8 million</td><td rowspan="5">Up to €120,202 of taxable profits</td></tr>
<tr><td>2006</td></tr>
<tr><td>2007</td><td>32.5%</td><td rowspan="6">25%</td></tr>
<tr><td>2008</td><td rowspan="5">30%</td></tr>
<tr><td>2009</td></tr>
<tr><td>2010</td><td rowspan="3">€10 million</td><td rowspan="3">Up to €300,000 of tax. profits</td></tr>
<tr><td>2011</td></tr>
<tr><td>2012</td></tr>
</table>

Source: the applicable laws are: Law 43/1995 (Article 122), Law 6/2000 (Article 122), Law 24/2001 (Article 122), Law 4/2004 (Article 108), Law 2/2004 of the *Presupuestos Generales del Estado* (Annual Government Budget Law, Article 108), and Real Decreto Ley 13/2010 (Royal Law Decree, Article 108).

Table 2: Bunching Estimations

	b_{NF}	b_F	B	H	y_{lb}	y_{ub}	N
Pooled data							
1999-2007	.082***	.428***	3394.3	3447.9	5.50	6.54	290,711
	(.005)	(.048)					
Annual data							
1999	.1438***	.721***	385.8	378.0	5.50	6.94	20,397
	(.014)	(.180)					
2000	.116***	.549***	382.6	367.2	5.50	6.70	23,525
	(.013)	(.120)					
2001	.082***	.534***	305.9	298.5	5.50	6.70	26,466
	(.013)	(.214)					
2002	.054***	.252***	224.5	217.3	5.50	6.30	29,086
	(.008)	(.053)					
2003	.070***	.323***	314.5	301.1	5.50	6.38	31,629
	(.009)	(.067)					
2004	.088***	.422***	429.5	408.6	5.50	6.50	34,540
	(.009)	(.077)					
2005	.074***	.364***	403.1	415.4	5.50	6.46	38,182
	(.008)	(.062)					
2006	.084***	.406***	530.7	536.3	5.50	6.50	43,538
	(.008)	(.065)					
2007	.069***	.345***	435.2	434.7	5.50	6.42	43,348
	(.009)	(.073)					

Note: b_{NF} and b_F are the bunching intensity parameters assuming no frictions and frictions, respectively. B is the number of firms above the counterfactual density of revenue in the range $y \in (y_{lb}, y^{LTU})$, where y is revenue, y_{lb} is the lower bound of the excluded region (used to construct the counterfactual) and y^{LTU} is the LTU threshold of €6 million. H is the missing number of firms below the counterfactual density in the range $y \in (y^{LTU}, y_{ub})$, where y_{ub} is the upper bound of the excluded region. Finally, N is the number of observations included in the estimations, i.e. the number of firms with revenue $y \in$ (€3,€9) million in each year. Significance levels: *** = 1%, ** = 5%, and * = 10%.

Table 3: Sensitivity Analysis for the Bunching Estimators

Excluded region		Point estimates	
Lower bound: y_{lb}	Upper bound: y_{ub}	No frictions: b_{NF}	Frictions: b_F
5.30	6.46	0.075***	0.368***
5.40	6.50	0.079***	0.403***
5.50	6.54	0.082***	0.428***
5.60	6.50	0.082***	0.420***
5.70	6.46	0.081***	0.394***

Note: This table shows the sensitivity of the bunching estimators to different assumptions on the excluded region used to estimate the counterfactual. We arbitrarily fix different values of y_{lb} in the first column and then obtain the corresponding value of y_{ub} and the point estimates for the bunching estimators $\widehat{b}_{NF}$ and $\widehat{b}_F$. These estimations use data only for the period 1999-2007. Significance levels: *** = 1%, ** = 5%, and * = 10%.

Table 4: Heterogeneity of the Response Across Groups

	No Frictions: b_{NF}	Frictions: b_F
By Number of Employees		
$1-25$ employees	.078***	.309***
	(.006)	(.040)
$26-40$ employees	.084***	.462***
	(.006)	(.069)
$41-50$ employees	.128***	1.167
	(.013)	(.808)
More than 50 employees	.078***	.745***
	(004)	(.116)
By Fixed Assets		
$0-0.6$ million Euros	.089***	.352***
	(.006)	(.042)
$0.6-2.4$ million Euros	.097***	.638***
	(.005)	(.084)
$2.4-5$ million Euros	.084***	.528***
	(.006)	(.075)
More than 5 million Euros	.019***	.209***
	(.006)	(.098)
By Organizational Form		
SA (Corporation)	.084***	.585***
	(.004)	(.066)
SL (L.L.C.)	.088***	.381***
	(.005)	(.039)
By Revenue Trend		
Growing	.106***	.524***
$(\overline{y}_t > \overline{y}_{t-1})$	(.004)	(.040)
Shrinking	.015	.940
$(\overline{y}_t < \overline{y}_{t-1})$	(.014)	(1.674)

Note: this table reports the bunching intensity estimates for "no frictions" and "frictions" ($\widehat{b}_{NF}$ and $\widehat{b}_F$) for different subsamples of firms. The subsample are defined by number of employees, by the level of fixed assets, by the type of organizational form, and by the firms' growing trends. In the latter case, $\overline{y}_t$ stands for reported revenue in year t. These estimates are obtained using data only for the period 1999-2007. Significance levels: *** = 1%, ** = 5%, and * = 10%.

Table 5: Bunching Persistence Over Time: Regression Results

	Dependent Variable: $Pr\,[bin\,(y_t) = bin\,(y_{t+h})]$							
	h=1	*h=2*	*h=3*	*h=4*	*h=5*	*h=6*	*h=7*	*h=8*
BunchBin	.063***	.032***	.021***	.017***	.010***	.009***	.005*	.003
	(.003)	(.003)	(.002)	(.002)	(.002)	(.002)	(.003)	(.003)
Constant	.084***	.055***	.042***	.036***	.032***	.027***	.017***	.013***
	(.002)	(.002)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Year FE	yes	yes	yes	yes	yes	yes	yes	yes
Obs.	668,943	590,124	519,773	444,343	369,329	300,291	238,778	182,878
Clusters	141,589	129,900	120,928	109,704	97,010	85,021	74,545	64,103
R^2	0.023	0.015	0.011	0.008	0.008	0.007	0.007	0.007

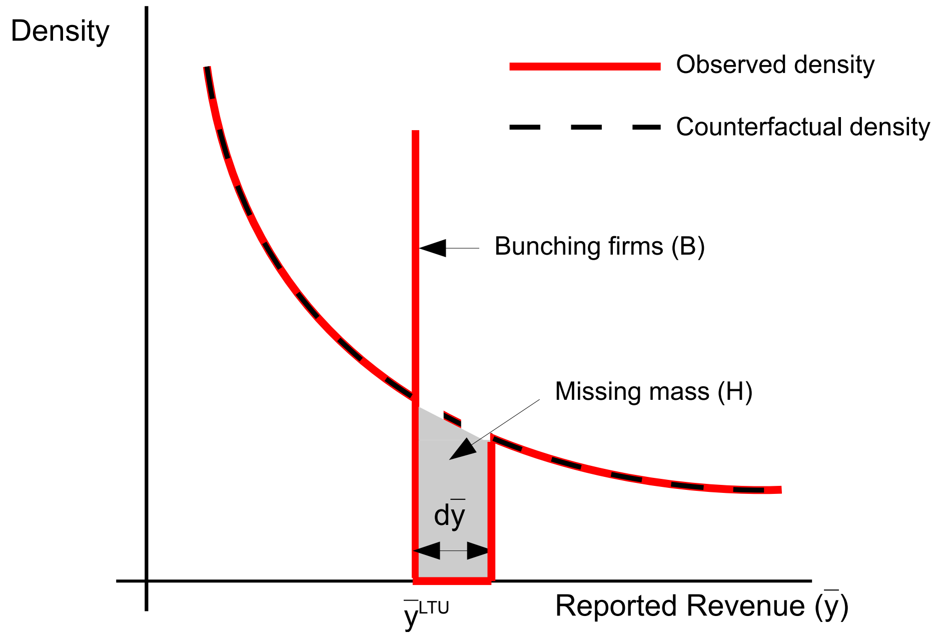
Note: this table reports coefficients from the following regression equation:

$$Prob\,[bin\,(y_{it}) = bin\,(y_{i,t+h})] = \alpha + \beta BunchBin_{it} + y_{it} + y_{it}^2 + \lambda_t + \varepsilon_{it},$$

where y_{it} is reported revenue by firm i in year t, the left-hand side variable is the fraction of firms that report revenues in the same bin in years t and $t+h$, $BunchBin_{it}$ takes value one if $y_{it} \in (5.75, 6]$, and λ_t denotes a vector of year dummies. We add a quadratic polynomial in current reported revenue as a way to control for the counterfactual probability that firms remain in a given revenue bin. Instead of using revenue levels, we use the distance to the notch so that the constant term α can be interpreted as the fraction of firms near the notch expected to remain at their current revenue level h years from now. These estimations use data for the full period 1999-2011.

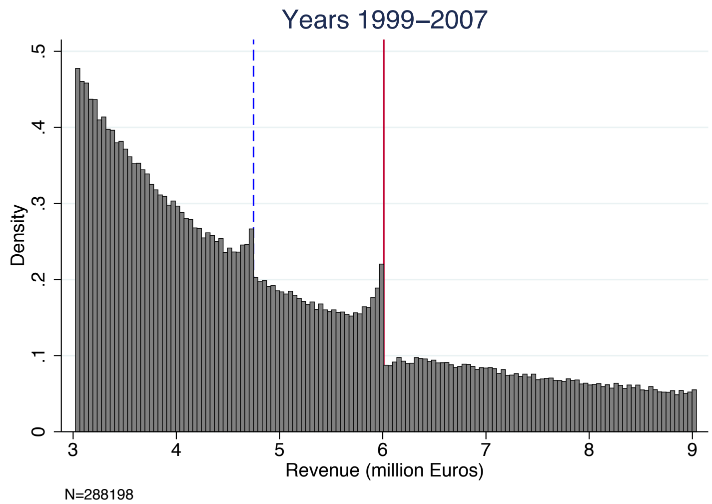
Figures

Figure 1: Theoretical Revenue Distribution



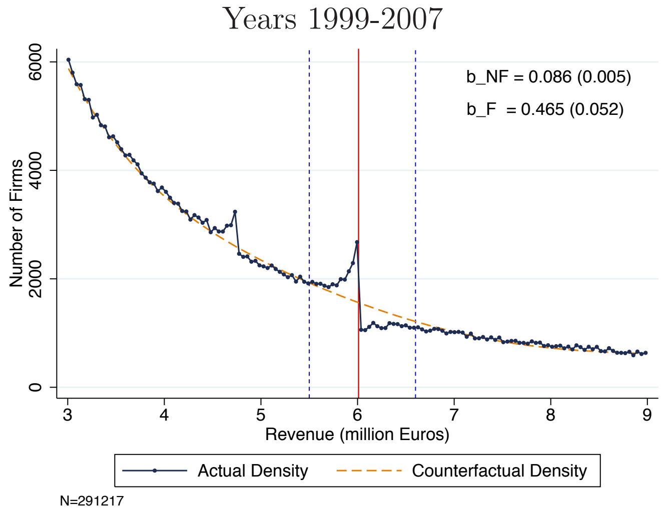
Note: this figure depicts the theoretical revenue distribution before and after the introduction of the Large Taxpayers' Unit (LTU). In the benchmark scenario, all firms face the same probability of detection and the distribution of revenue is smoothly decreasing as depicted by the dashed (black) line. When the LTU is introduced, firms reporting revenue above y^{LTU} face a higher enforcement intensity. A group of firms in an interval above y^{LTU} respond to the new policy by underreporting more of their revenue to report exactly $\bar{y} = y^{LTU}$. This generates a spike at the threshold (with excess mass B), and an area of missing mass (H) to the right of the threshold, as depicted by the solid (red) line. Notice that this plot assumes that there are no optimization frictions, so all firms can immediately respond to fiscal incentives. Thus, there are no firms in the interval of length $d\overline{y}$ above the threshold.

Figure 2: Operating Revenue Distribution



Note: the histograms pool data for the period 1999-2007. The dashed (blue) line indicates the External Audit threshold, set at €4.75 million. The solid (red) line indicates the LTU threshold, set at €6 million. The bins are exactly €42,070 wide and are defined such that no bin contains data both to the left and to the right of each threshold.

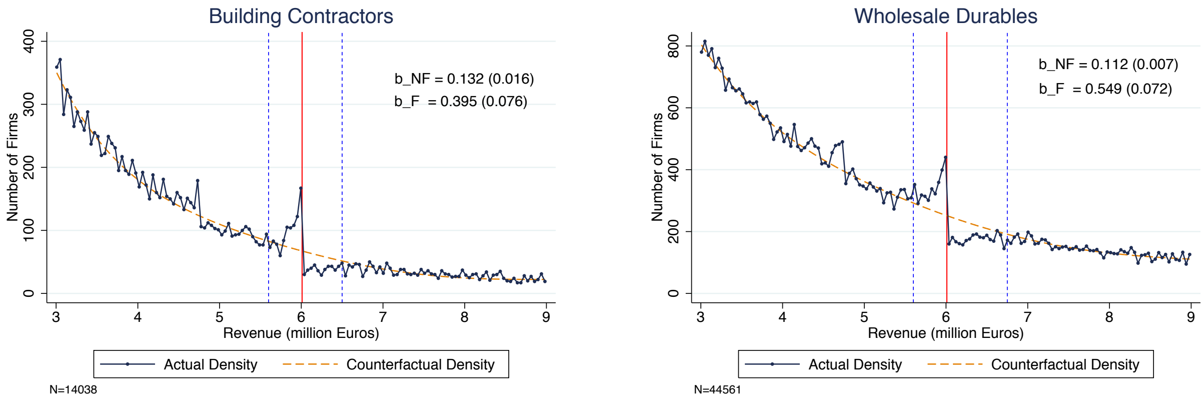
Figure 3: Counterfactual Distribution & Bunching Estimates



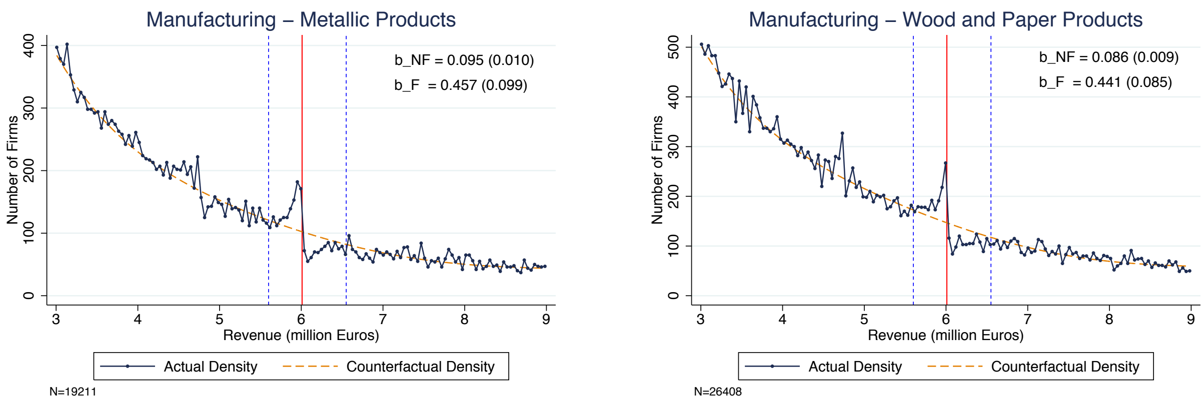
Note: these graphs shows the reported distribution of revenue (dots connected by solid blue line) and the estimated counterfactual (orange dashed curve) for the period 1999-2007. The data for the true distribution are exactly the same as that used to construct the histogram in Figure 2. The vertical dotted blue lines indicated the bounds of the excluded region (y_{lb} and y_{ub}) chosen for the estimation of the counterfactual. To determine the value of y_{ub}, we fit a polynomial regression to the true density multiple times, starting with $y_{ub} \approx y^{LTU}$ and increasing the value in small steps until we reach a point where the bunching mass (B) equals the missing mass (H). This way, the area under the counterfactual density is the same as under the true density. "b_NF" denotes the estimate of bunching intensity derived under the assumption of no optimization frictions (b_{NF}), and "b_F" denotes the estimate that takes into account the existence of frictions (b_F).

Figure 4: Revenue Distribution by Sector of Activity

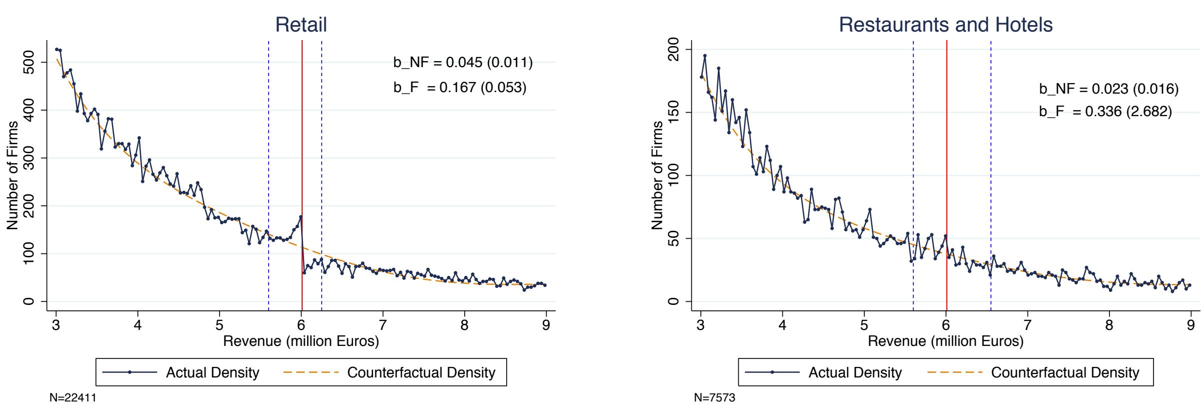
(a) High Bunching Sectors



(b) Medium Bunching Sectors

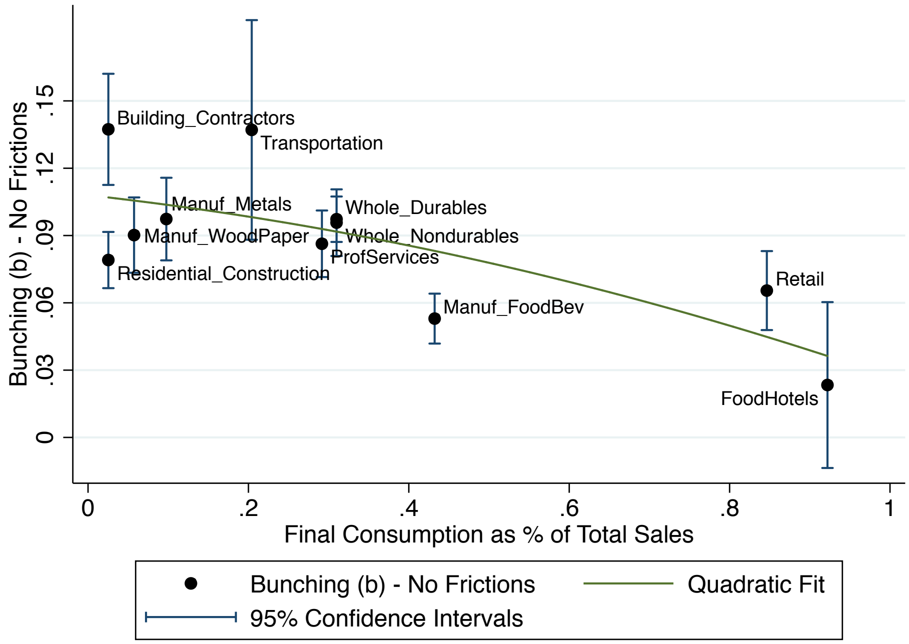


(c) Low Bunching Sectors



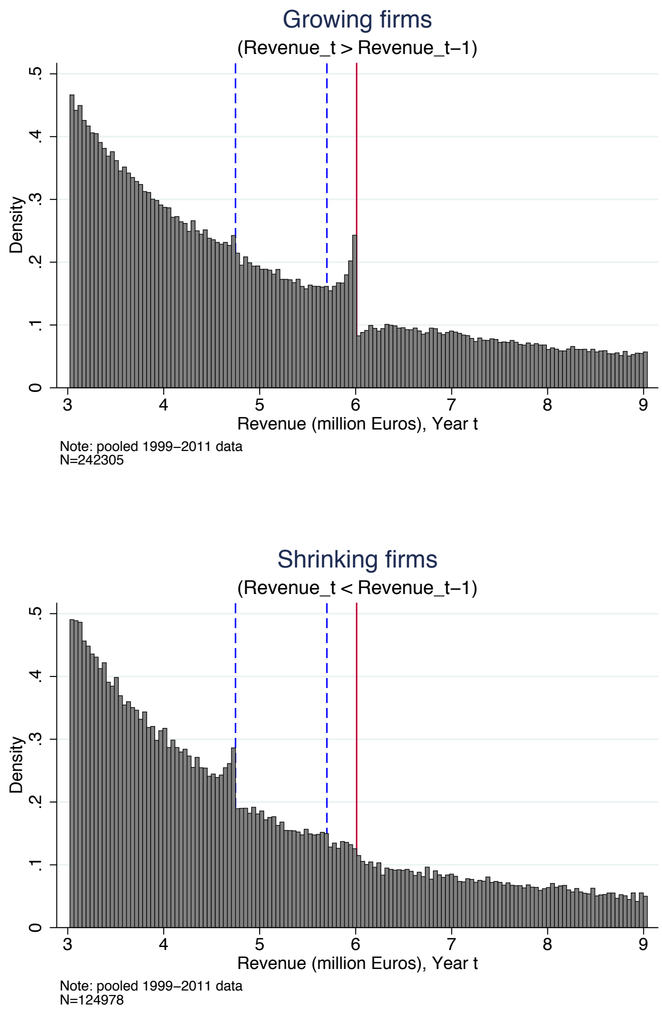
Note: these graphs show the actual and counterfactual revenue distributions for selected sectors (six out of a total of 12 sectors defined). The counterfactual distribution is constructed in each case as explained in the note to Figure 3. Only data for the period 1999-2007 is used.

Figure 5: Bunching Response by Scope of Evasion



Note: the bunching measure $\widehat{b}_{NF}$ is calculated for each sector as explained in Section 4 in the main text. Final consumption as a share of total sales in a sector is calculated using the year 2000 input-output tables for the Spanish economy, published by the National Statistics Institute (INE). The assumption underlying this figure is thst selling to final consumers makes underreporting much easier because the VAT self-enforcing mechanism breaks down at the last stage of the value chain.

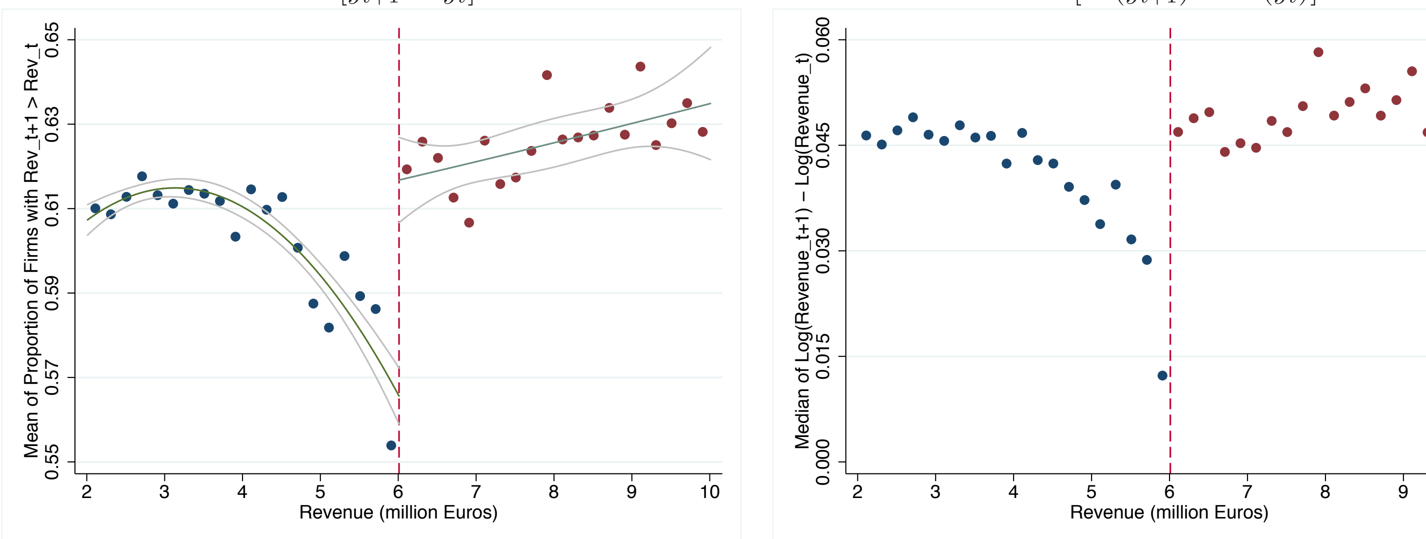
Figure 6: Growing vs. Shrinking Firms



Note: these graphs show annual revenue distributions for two subsamples of firms: those that are growing and those that are shrinking. A firm is defined as growing if its reported revenue in year t is higher than in year $t-1$, i.e. $\overline{y}_t > \overline{y}_{t-1}$. A firm is defined as shrinking if its reported revenue in year t is lower than in year $t-1$, i.e. $\overline{y}_t < \overline{y}_{t-1}$. The dashed (blue) line indicates the External Audit threshold, set at €4.75 for 1999-2007 and €5.7 million for 2008-2011. The solid (red) line indicates the LTU threshold, set at €6 million for both periods (and fixed in nominal terms). The bins are exactly €42,070 wide and are defined such that no bin contains data both to the left and to the right of each threshold. The graphs pool data for the entire period 1999-2011, but the pattern is consistent in both 1999-2007 and 2008-2011.

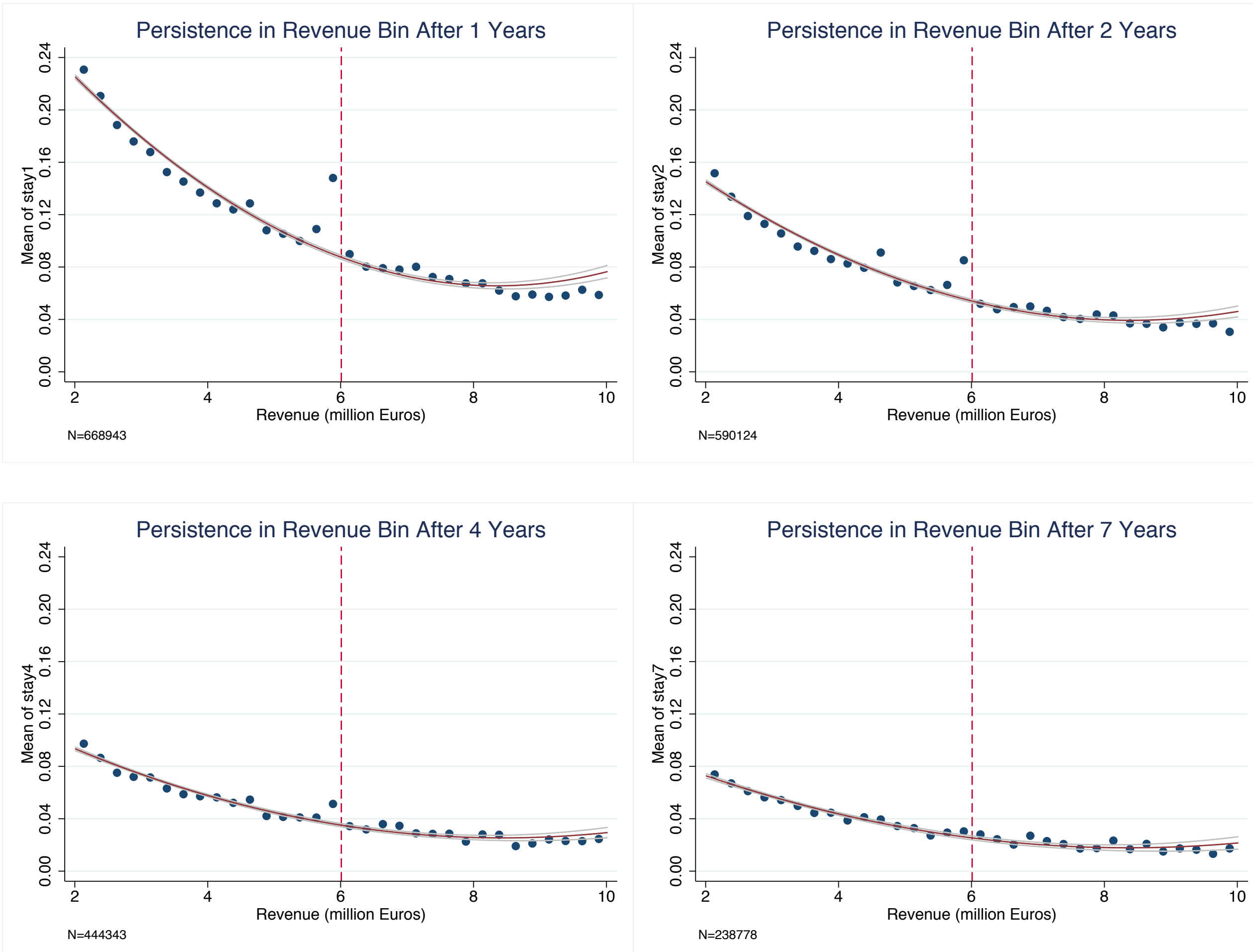
Figure 7: Dynamic Behavior around the LTU Threshold

Probability of Positive Revenue Growth	Median Revenue Growth
Prob $[y_{t+1} > y_t]$	Median $[\ln(y_{t+1}) - \ln(y_t)]$



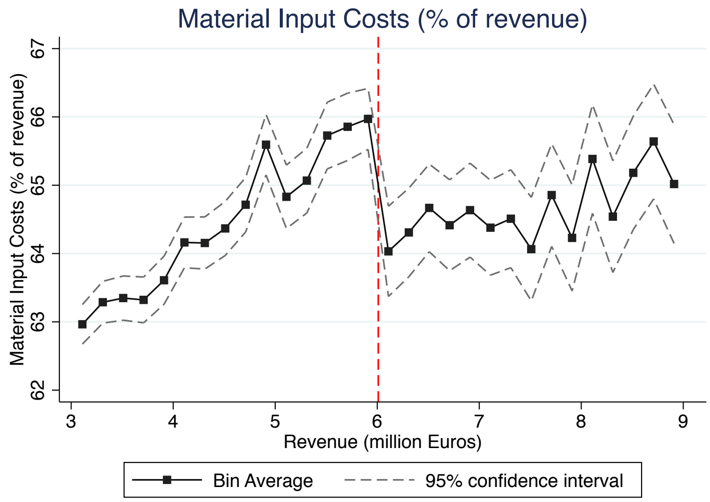
Note: the left panel shows the probability that firms report higher revenues in year $t+1$ than t, against reported revenue in year t. The right panel shows median growth in reported revenue experienced by firms in year $t+1$, against reported revenue in year t. Data are divided in bins of €200,000. The dots depict bin averages, while the solid lines in the left panel are quadratic fits estimated separately on either side of the threshold. The dashed lines indicate 95% confidence intervals around the quadratic fits.

Figure 8: Bunching Persistence



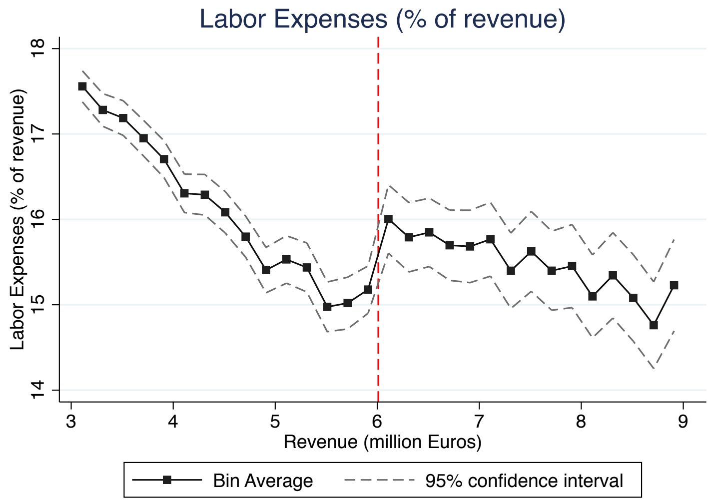
Note: these graphs show the degree of bunching persistence. Data are divided in €250,000 bins such that no bin includes firm both to the left and to the right of the threshold, marked by the dashed (red) vertical line. The blue dots indicate the proportion of firms who reported revenues within that bin both in year t and year $t+h$, where h is the number of years. The solid red curve is a quadratic fit of the bin averages, excluding the bunching bin, i.e. the interval of reported revenue $y \in (5.75, 6]$.

Figure 9: Material Input Costs



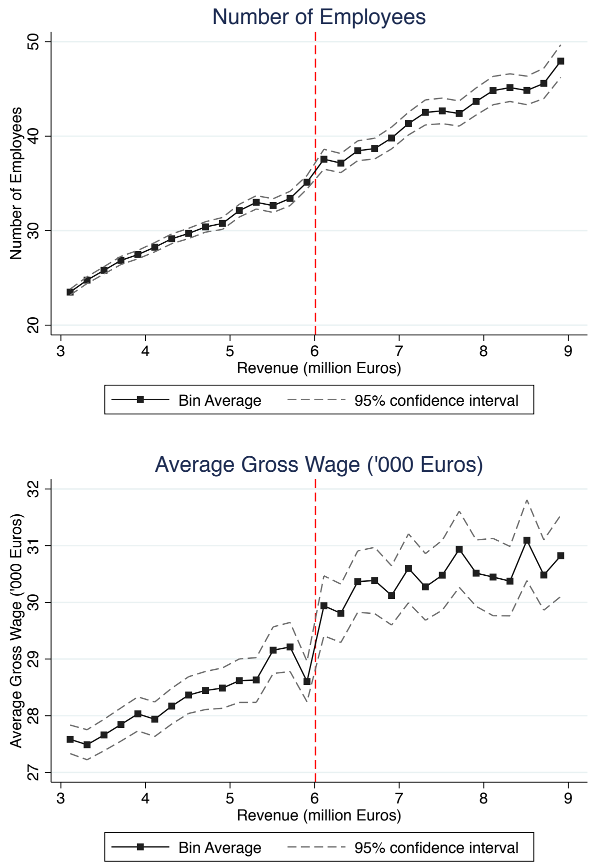
Note: this graph shows average material input costs over revenue by ranges of reported revenuefor the period 1999-2007. Data are divided in €250,000 bins such that no bin includes firms both to the left and to the right of the LTU threshold, which is marked by the dashed (red) vertical line. The dots depict bin averages and the dashed lines are 95% confidence intervals. We trim outliers in the data by dropping the top and bottom 0.5% of observations for the outcome variable.

Figure 10: Labor Input Costs



Note: this graph shows average labor input costs over revenue by ranges of reported revenue for the period 1999-2007. The data are divided in €250,000 bins such that no bin includes firms both to the left and to the right of the LTU threshold, which is marked by the dashed (red) vertical line. The dots depict bin averages and the dashed lines are 95% confidence intervals. We trim outliers in the data by dropping the top and bottom 0.5% of observations for the outcome variable.

Figure 11: Number of Employees and Average Wages



Note: these graphs show average number of employees (top panel) and average gross wages (bottom panel) by ranges of reported revenue for the period 1999-2007. The data are divided in €250,000 bins such that no bin includes firms both to the left and to the right of the LTU threshold, which is marked by the dashed (red) vertical line. The dots depict bin averages and the dashed lines are 95% confidence intervals. We trim outliers in the data by dropping the top and bottom 0.5% of observations for the outcome variable.

Appendix: Not Intended for Publication

A Model Extensions

The model in Section 2 assumes that there is only one production input, m, and one tax, the corporate income tax. In that model, firms can only respond to tax enforcement regulations by modifying their reported revenue through changes in output (real response) or by misreporting their revenue to evade the corporate income tax (evasion response). Here, we enrich this theoretical framework in several ways. In a first extension, we allow firms to also misreport their input costs, besides misreporting their sales. Overreporting input costs can be advantageous because it lowers reported profits and therefore the overall tax bill if the firm is not detected. Once input misreporting has been added to the model, a natural step is to include in the model the value added tax (VAT), which creates additional incentives to overreport inputs. The second extension is to consider a production function with two inputs: labor and materials. Considering these two inputs is interesting because the tax incentives associated with each of these inputs are different. It is also convenient because the dataset we use includes accurate measures of firms' total expenditures on both of them.

We begin by setting up the two model extensions and explaining how the setup of the firm's maximization problem changes firm incentives. Then, we summarize the testable predictions generated by the two models, focusing in particular on the behavior of firms around the LTU threshold.

A.1 Model with Input Overreporting

Consider a situation in which firms have two ways of manipulating their taxable income: they can underreport their sales and overreport their input costs. Both activities lead to a reduction in reported profits, and hence to a lower tax payment if not detected. Let $u = y - \overline{y}$ denote the amount of sales underreported and let $v = \overline{m} - m$ be the amount of inputs overreported. Notice that $u \geq 0$ and $v \geq 0$ by construction, because it is never beneficial for firms to overreport sales or underreport inputs (they would pay higher taxes without receiving any additional benefit). In this setting, true and reported before-tax profits are given by:

$$\Pi = p\psi f(m) - cm \tag{20}$$

$$\overline{\Pi} = p[\psi f(m) - u] - c[m + v] \tag{21}$$

The tax authority monitors firms to detect tax evasion behavior and tax audits always uncover the full amount of taxes evaded. As before, let θ be the penalty rate applied to the amount of evasion detected and let δ denote that probability of evasion detection, which is a convex function of total firm output, sales underreporting and input overreporting. Formally, we have $\delta = \delta(m, u, v)$ with:

$$\begin{aligned} \partial\delta/\partial m > 0 \qquad & \partial^2\delta/\partial m^2 > 0 \\ \partial\delta/\partial u > 0 \qquad & \partial^2\delta/\partial u^2 > 0 \\ \partial\delta/\partial v > 0 \qquad & \partial^2\delta/\partial v^2 > 0 \end{aligned}$$

Firms choose material inputs, m, underreported sales, u, and overreported inputs, v, to maximize expected profits, subject to the technological constraint, prices given by competitive markets and tax enforcement policies. Expected profits are then given by:

$$\mathbb{E}\Pi = \Pi - t\overline{\Pi} - \delta t\,(1+\theta)\left[\Pi - \overline{\Pi}\right] \tag{22}$$

where the difference between true and reported profits can be written as follows:

$$\Pi - \overline{\Pi} = pu + cv \geq 0 \tag{23}$$

The first order conditions for an interior optimum are given by:

$$(1-t)\,p\psi f_m(m^*) = (1-t)\,c + t\,(1+\theta)\,\frac{\partial\delta}{\partial m}\,[pu + cv] \tag{24}$$

$$tc = t\,(1+\theta)\left[\frac{\partial\delta}{\partial v}\,[pu + cv*] + \delta c\right] \tag{25}$$

$$tp = t\,(1+\theta)\left[\frac{\partial\delta}{\partial u}\,[pu * + cv] + \delta p\right] \tag{26}$$

Equation (24) shows that when either $u > 0$ or $v > 0$, optimal input purchases are affected by the possibility of tax evasion. Firms buy less inputs than in an economy without tax evasion because the possibility of evasion raises the marginal cost of acquiring inputs. The next two equations characterize the optimal misreporting choices: input overreporting and sales underreporting. According to condition (25), firms equalize the marginal tax savings of overreporting input costs to the marginal expected payment if detected. According to condition (26), firms equalize that marginal tax savings of underreporting sales to the marginal expected payment if detected.

Solving (25) and (26) for v^* and u^* and introducing them into (24), we obtain the following expressions:

$$\frac{\partial \delta / \partial v}{\partial \delta / \partial m} = \frac{tc\left[1-\delta\left(1+\theta\right)\right]}{\left(1-t\right)\left[p\psi f_m\left(m^*\right)-c\right]} \tag{27}$$

$$\frac{\partial \delta / \partial u}{\partial \delta / \partial m} = \frac{tp\left[1-\delta\left(1+\theta\right)\right]}{\left(1-t\right)\left[p\psi f_m\left(m^*\right)-c\right]} \tag{28}$$

where $tc\left[1-\delta\left(1+\theta\right)\right]$ and $tp\left[1-\delta\left(1+\theta\right)\right]$ are the expected marginal returns of input overreporting and sales underreporting, respectively. Condition (27) indicates that the relative increase in the detection probability due to a higher use of inputs (more production) and a higher amount of input overreporting (more evasion) must be equal to the relative return between evading taxes and acquiring inputs. Condition (28) has a similar interpretation, but with the additional evasion coming from revenue underreporting.

The key intuition obtained from this extension of the model is that firms can evade taxes through two symmetric channels, sales underreporting and input overreporting. In the first order conditions, one additional euro of sales underreporting leads to the same tax evasion as one additional euro of inputs overreporting. However, this symmetry breaks down when firms are close to the LTU threshold, because sales underreporting can determine whether the firm faces high or low enforcement, whereas the amount of inputs reported does not. Because of the discontinuity in enforcement, bunching firms are at a corner solution, so their optimum is not characterized by the first order conditions.

Introducing the value-added tax

To make the model more applicable to the context under study, we introduce the value-added tax (VAT) into the model. The VAT is designed to be a tax on consumption, with firms playing the role of fiscal intermediaries that help in the process of tax collection. VAT is charged on every business transaction, regardless of whether a sale is made to a final consumer or to a firm as an intermediate input. At the end of the fiscal period (usually a month or a quarter), firms calculate all the VAT they have charged on their sales and all the VAT they have paid on their inputs and remit to the tax agency the difference between the two. If the balance is negative, then the firm receives a reimbursement. Absent other distortions and assuming that firms report their transactions truthfully, the VAT does not distort productive efficiency. To show this in a simple way, it is convenient to introduce some additional notation and work only with monetary amounts. Let $Y = p\psi f\left(m\right)$ denote total revenue from sales and let $M = cm$ denote total input costs. True after-tax profits under truthful reporting with both a corporate income tax and a VAT are:

$$
\begin{aligned}
\Pi &= \left(1 - t^{cit}\right)\left[\left(1 + t^{vat}\right)(Y - M) - t^{vat}(Y - M)\right] \\
&= \left(1 - t^{cit}\right)[Y - M]
\end{aligned} \tag{29}
$$

where t^{cit} is the corporate income tax rate and t^{vat} is the value-added tax rate. When misreporting of either revenues or input costs is allowed, the neutrality of VAT is broken. By engaging in misreporting, the firm takes advantage of its role as a fiscal intermediary and keeps some resources that should have been transferred to the government as part of the VAT collection process. Actual profits obtained by the firm when misreporting is allowed and the firm is not detected (ND) are:

$$
\begin{aligned}
\Pi^{ND} &= \left(1 - t^{cit}\right)\left\{\left(1 + t^{vat}\right)\left(\overline{Y} - \overline{M}\right) - t^{vat}\left(\overline{Y} - \overline{M}\right)\right\} + (U + V) \\
&= \left(1 - t^{cit}\right)[Y - M] + \left(t^{cit} + t^{vat}\right)(U + V)
\end{aligned} \tag{30}
$$

As in previous versions of the model, the tax agency detects the evasion behavior with some probability δ and applies a penalty rate θ. Profits obtained by the firm when evasion is detected are:

$$
\Pi^{D} = \left(1 - t^{cit}\right)[Y - M] - (1 + \theta)\left(t^{cit} + t^{vat}\right)(U + V) \tag{31}
$$

Therefore, the expression for expected profits is similar to previous ones, with the additional incentive to evade due to the VAT:

$$
\mathbb{E}\Pi = \left(1 - t^{cit}\right)[Y - M] + [1 - \delta(1 + \theta)]\left(t^{cit} + t^{vat}\right)(U + V) \tag{32}
$$

With equation (32) as an objective function, one can derive the solution of the model as before by simply defining total evasion E as the sum of sales underreporting and input overreporting: $E = U + V$.

A.2 Model with Two Production Inputs

Up to this point, all the models we have worked with allow firms to use only one production input. This restriction is clearly unrealistic, because in practice firms use a variety of inputs in the production process. Interestingly for the context of this paper, the tax incentives are not the same for different production inputs. In this extension, we consider a model with two production inputs: materials and labor. To match the aggregated definitions of inputs in the Amadeus data, we consider material inputs to include both raw materials and external services used for production. The measure of labor inputs is the total wage bill of the firm.

Let $N = wn$ denote the total wage bill, where w is the market wage rate and n is the number of employees. Also, let t^{prt} be the statutory payroll tax rate. As explained above, Spanish Law assigns part of the payroll tax to the employer and part to the employee. Ultimately, the incidence of the tax is an empirical question unrelated to the statutory taxes, so we abstract from this and use a single tax rate that includes both the employer's and the employee's share. True after-tax profits are then given by:

$$\begin{aligned}\Pi &= \left(1-t^{cit}\right)\left\{\left(1+t^{vat}\right)[Y-M]-t^{vat}[Y-M]-\left(1+t^{prt}\right)N\right\}\\ &= \left(1-t^{cit}\right)\left\{[Y-M]-\left(1+t^{prt}\right)N\right\}\end{aligned} \tag{33}$$

Equation (33) yields some standard results. First, neither the corporate income tax nor the VAT distort production decisions when there truthful reporting. The payroll does increase the marginal cost of labor, so it leads to a suboptimally low employment level in equilibrium.

We now allow for the possibility of misreporting labor costs, letting $Z = N - \overline{N}$ denote underreported labor costs. In this case, after-tax profits when evasion is not detected are given by:

$$\begin{aligned}\Pi^{ND} &= \left(1-t^{cit}\right)\left\{\left(1+t^{vat}\right)\left[\overline{Y}-\overline{M}\right]-t^{vat}\left[\overline{Y}-\overline{M}\right]-\left(1+t^{prt}\right)\overline{N}\right\}+(U+V-Z)\\ &= \left(1-t^{cit}\right)\{Y-M-N\}+\left(t^{cit}+t^{vat}\right)(U+V)+\left[t^{prt}-t^{cit}\left(1+t^{prt}\right)\right]Z\end{aligned} \tag{34}$$

Profits if detected by the tax authorities are derived in a similar way:

$$\begin{aligned}\Pi^{D} = {} & \left(1-t^{cit}\right)\{Y-M-N\}-\\ & -(1+\theta)\left\{\left(t^{cit}+t^{vat}\right)(U+V)+\left[t^{prt}-t^{cit}\left(1+t^{prt}\right)\right]Z\right\}\end{aligned} \tag{35}$$

Finally, we obtain the usual expression for expected profits that firms try to maximize:

$$\begin{aligned}\mathbb{E}\Pi = {} & \left(1-t^{cit}\right)[Y-M-N]+\\ & +[1-\delta(1+\theta)]\left\{\left(t^{cit}+t^{vat}\right)(U+V)+\left[t^{prt}-t^{cit}\left(1+t^{prt}\right)\right]Z\right\}\end{aligned} \tag{36}$$

Material inputs are deductible under the value added tax (VAT) and the corporate income tax (CIT). Hence, overreporting material inputs unambiguously lowers the amount of VAT and CIT remitted to the government (if not detected). There is widespread anecdotal evidence of firms overreporting materials inputs in Spain. For example, firms tend to include personal expenditures of CEOs and senior management into the company books. There are multiple reports of this practice with durable goods such as automobiles (which are really intended for personal use) and also with large social events such as weddings (reported as company events).

Labor inputs cannot be deducted from the VAT, but they are instead taxed through the payroll tax. Underreporting labor inputs lowers the amount of payroll tax remitted, but it increases tax liability on the CIT. Therefore, the incentive to over- or underreport labor inputs depends on the relative marginal tax rates of the payroll tax and the CIT. Specifically, underreporting is advantageous as long as $t^{prt} < t^{cit}(1 + t^{prt})$. In the period we study, the tax rates were $t^{prt} = 38\%$ and $t^{cit} = 35\%$ (reduced to 30% in 2007). Applying these rates to the formula yields a small incentive to overreport labor costs.[40] However, there are two important factors in favor of underreporting of labor costs that this model does not capture: potential collusion with workers and downward wage rigidities. We explain how these two factors work below.

If wages are underreported, employees face a lower personal income tax than they would with truthful reporting. Even though they also lose some potential future benefits like higher pensions and unemployment insurance payments, those are small compared to the savings from evading income taxes today. Hence, we argue that there are strong incentives for wage earners to collude with their employers to underreport wages. Evidence on this practice among firms is widespread in Spain[41] and other countries.[42]

Downward nominal wage rigidity provides an additional reason for firms to underreport wages. In good years, firms would like to raise their employees' wages, but they know that in bad years it will be extremely difficult to lower in a symmetric way due to the power of unions and an inflexible collective bargaining system. In this context, firms can use the

[40] With $t^{cit} = 35\%$, the marginal return on each euro of labor costs underreported would be $0.38 - 0.35\,(1.38) = -0.10$. With the lower tax rate of 30%, the return gets closer to zero: $0.38 - 0.30 * (1.38) = -0.034$. A marginal corporate income tax rate of 27.3% would make firms indifferent between over- and underreporting labor costs. After the CIT reform in 2007, $t^{cit} = 25\%$ for firms with reported revenue below €8 million.

[41] For example, there are open judicial investigations on the political party currently in power at the national level and on the vice-president of the National Employers Federation for paying salary "complements" in cash.

[42] Kumler, Verhoogen and Frias (2012) provide evidence of wage underreporting in Mexico, where many firms report payments barely above the minimum wage to evade payroll taxes, while average and median wages reported in household surveys are two or three times above the reported amounts.

cash revenues obtained through unreported sales to give wage "bonuses" to their employees in good times, and pay them only their official salary in bad years.

A.3 Theoretical Predictions

In deriving testable predictions from the models presented above, we make the assumption that a random subset of firms is affected by optimization frictions. This means that these firms are not able to respond to the incentives around the LTU threshold by misreporting their activities. Even though we have not modeled optimization frictions explicitly up to this point, we know from the empirical revenue distribution shown in Section 4 that they are substantial because many firms report revenues just above the LTU threshold.

Understanding the use of production inputs by firms above and below the threshold can shed light on what type of behavioral response is dominant. We use the model extensions presented above to derive predictions of how the reported input costs would look like under several scenarios. Assuming that there are optimization frictions is necessary for this predictions to make sense, because otherwise there would be no firms at all just above the threshold. Another assumption, implicit in our production function, is that the ratios of inputs over revenue are constant for all levels of revenue. This is not exactly true in the data, but it is a good approximation for short intervals around the LTU threshold.

Model with One Input

In the model with one production input and two taxes (corporate income and VAT), the predictions are straightforward. If the bunching response is fully real, concavity of the production function $f(m)$ implies that firms just below the threshold must use (and hence report) lower input costs on average. This is because bunching firms are more productive than those that would have been below the notch even in the absence of the LTU. The firms that do not respond to the incentives because of optimization frictions and remain just above the notch report a higher input/revenue ratio than the bunching firms. To sum up, if the bunching response is through a decline in production, we should observe an upward jump in the input/revenue ratio at the LTU threshold.

If the response is fully due to evasion, bunching firms obtain revenues above €6 million but report a smaller amount. Assuming bunching firms report inputs truthfully, the reported ratio of inputs over revenue would be relatively high. If they also overreport input costs (to take advantage of the lower enforcement intensity), the reported ratio will be even higher.

Thus, the model with sales underreporting and input overreporting predicts a downward jump in the input/revenue ratio at the LTU threshold.

Model with Two Inputs

In the model with two inputs (materials and labor) and two taxes (corporate income and VAT), the predictions for the ratio of material inputs over total revenue are the same as for a single input. If the response is fully real, we should observe an upward jump in the ratio at the LTU threshold, because bunching firms are relatively more productive. If the response is fully due to revenue underreporting, we should observe a downward jump in the ratio, larger if bunching firms also overreport their materials.

The predictions for the ratio of labor costs over revenue are less clear-cut. Under a fully real response, the prediction is the same as for materials: we should observe an upward jump in the ratio of labor costs over revenue at the LTU threshold. If the bunching response is fully due to evasion and there is no misreporting of labor costs, then we would expect a downward jump in the ratio. Once we allow for labor cost misreporting, the incentives depend on the corporate income and payroll tax rates as explained above, apart from the other incentives for underreporting discussed (potential for collusion and downward nominal wage rigidity). Thus, the prediction in the latter case is ambiguous.

B Institutional Background and Data: Additional Details

B.1 Overview of the Spanish Tax System

The Spanish tax system rests on four main taxes: the payroll tax (PRT),[43] the individual income tax (IIT), the value-added tax (VAT) and the corporate income tax (CIT). The payroll tax accounted for a stable 33% of all tax revenues in the period 1999-2011, followed by the IIT with 22%, the VAT with 19%, and the CIT with 13% (the latter with wide fluctuations between 15% in boom years and 9% in recession years). The rest is collected through other indirect taxes and fees (IEF, 2011).

The top marginal tax rate on the individual income tax (IIT) in Spain was 48% in 1999-2002 and then lowered to 46% in 2003-2011. This rate is substantially higher rate

[43] In Spain these are known as Social Security Contributions (*"Cotizaciones Sociales"*), but the term can be confusing because Social Security includes multiple social protection programs, not just pensions and disability insurance as in the United States.

than the 35% (lowered to 30% in 2007) tax rate of the corporate income tax, which is 5 percentage points lower for firms under a revenue threshold that has changed over time. Thus, unlike in the US, high-income individuals have an incentive to shift taxable income from the IIT base to the CIT base to lower their tax liability. This seems to have led to the creation an abnormally large number of small firms in this period. The payroll tax rate was 38%, adding up the rates assigned to employers (31%) and employees (7%). Since wage negotiations usually focus on the net-of-tax wage, this is the most natural way of thinking about the overall payroll tax rate.[44] The general VAT rate in Spain during 1999-2010 was 16% (increased to 18% for 2010-11), with reduced rates for some goods and services.[45] The VAT is collected by firms at each production stage and then remitted to the government.

Penalties for tax evasion vary depending on the size of the infraction. If the amount evaded is above €120,000 (€90,000 prior to 2004), then the taxpayer faces criminal responsibility, whereas if it is below there can only be administrative penalties. There is a great deal of discretion regarding penalties, which according to law could go from 50% to 600% of the amount evaded, depending on the gravity of the offense. Fiscal crimes legally prescribe after four years, which in some cases limits the tax agency's ability to recover fiscal debts because the legal process is too slow.[46]

B.2 Data

As noted in the main text, small firms are under-represented in the Amadeus dataset. We assess the representativity of the sample by comparing the number of firms in Amadeus to the official statistics of the corporate income tax. The numbers on Table 6 confirm that small firms are under-represented in Amadeus (less than 50% compared to official statistics). On the other hand, more than 80% of firms with reported revenue between €3 and €9 million, which are the focus of our empirical analysis, are included in Amadeus.

Another limitation of the Amadeus data, common to the corporate tax literature, is that the financial statements do not provide an accurate measure of actual tax liability, because we do not observe the tax deductions applied by each firm to arrive at fiscal profit.[47] To know the exact tax liability, we would need administrative tax return data for all the major taxes,

[44] We do not get into the question of who bears the tax burden of the tax, since there is also a debate about the incidence of the corporate income tax itself.

[45] There were two reduced rates of 7 and 4% for items like staple foods, medicine and culture-related goods and services. Education and financial services were fully exempted from VAT.

[46] On the other hand, the tax inspector can request financial statements from the previous four years during an audit, and the company is legally obliged to provide them.

[47] The dataset does include a self-reported estimation of corporate income tax liability.

which is not available to researchers. Aggregate data published by the tax agency (AEAT, Several years) shows that the *effective* corporate tax rate paid by small and medium firms is higher than that of very large ones (25% vs. 22%), even though the statutory rate is higher for the latter group (30% vs. 35%). This indicates that tax deductions are of second-order importance for the size range we study. The information submitted to the Commercial Registry is essentially the starting point of the tax return, and the amount must match exactly.

B.2.1 Sectors of Activity: Definitions

TBD

C Empirical Results for the Recession Period, 2008-2011

C.1 Revenue Distribution

Figure 15 shows that the distribution of reported revenue is quite different for the recession period 2008-2011. First, the amount of bunching below the LTU threshold is substantially smaller than in the previous period, although still visible. Second, the External Audit threshold moved from €4.75 to €5.7 million, as indicated by the blue vertical dashed line. There is some bunching of firms at the new External Audit threshold, while the distribution is smooth over the old threshold. The fact that the two thresholds are much closer in this later period complicates the estimation, as explained below.

Figures 13 and 14 show the bunching pattern for each individual year in the periods 1999-2007 and 2008-2011, respectively. The distribution of reported revenue is remarkably stable and similar to the pooled data in the first period, with slightly noisier patterns given the smaller sample sizes. In the second period, the bunching response is consistently small every year, although 2009 stands out because there is no discernible bunching below the LTU threshold. That year the Spanish economy shrank by 3.7% and a very large share of firms faced negative (reported) revenue growth.

Appendix Tables

Table 6: Amadeus Dataset Compared to Official Statistics

		All Firms		€0-€3 million		€3-€10 million		€10+ million	
1999	Official	792,973		752,698		28,476		10,798	
	Amadeus	250,385	31.6%	218,429	29.0%	23,144	**81.3%**	8,812	81.6%
2000	Official	876,530		828,082		34,014		14,433	
	Amadeus	286,837	32.7%	249,401	30.1%	26,688	**78.5%**	10,748	74.5%
2001	Official	928,897		874,992		37,382		16,523	
	Amadeus	370,174	39.9%	328,040	37.5%	29,885	**79.9%**	12,249	74.1%
2002	Official	1,008,744		951,152		40,388		17,204	
	Amadeus	444,215	44.0%	398,015	41.8%	32,887	**81.4%**	13,312	77.4%
2003	Official	1,041,527		979,918		43,246		18,363	
	Amadeus	488,076	46.9%	437,670	44.7%	35,730	**82.6%**	14,676	79.9%
2004	Official	1,117,005		1,050,143		46,806		20,056	
	Amadeus	523,405	46.9%	468,128	44.6%	39,023	**83.4%**	16,252	81.0%
2005	Official	1,200,267		1,126,588		51,062		22,617	
	Amadeus	583,992	48.7%	522,679	46.4%	43,139	**84.5%**	18.168	80.3%
2006	Official	1,293,419		1,210,736		56,952		25,731	
	Amadeus	664,679	51.4%	594,443	49.1%	49,265	**86.5%**	20,966	81.5%
2007	Official	1,410,188		1,321,500		60,699		27,989	
	Amadeus	610,974	43.3%	539,977	40.9%	49,148	**81.0%**	21,843	78.0%
2008	Official	1,417,906		1,335,081		57,401		25424	
	Amadeus	656,511	46.3%	593,336	44.4%	43,247	**75.3%**	19,924	78.4%
2009	Official	1,414,877		1,347,188		46,113		21,576	
	Amadeus	576,576	40.8%	526,623	39.1%	33,490	**72.6%**	16,462	76.3%
2010	Amadeus	629,201	n/a	577,064	n/a	34,714	n/a	17,419	n/a
2011	Amadeus	503,120	n/a	462,488	n/a	27,138	n/a	13,489	n/a

Note: The percentages indicate the proportion of firms with complete revenue data in Amadeus compared to the number of firms that submitted a corporate income tax return that year. Official statistics are from several issues of *"Memoria de Administración Tributaria"*, an annual report published by the Spanish tax agency (AEAT, Several years). Official data for the years 2010 and 2011 are not yet publicly available. The Amadeus dataset is described in detail in section 3.2.

Table 7: Overview of the Spanish Tax System

	Top tax rate	*Share of tax revenue*
Social Security Contributions (SSC)	38%	33%
Individual Income Tax (IIT)	48% (46%)	22%
Value-Added Tax (VAT)	16%	19%
Corporate Income Tax (CIT)	35% (30%)	13%
Other indirect taxes and fees	-	13%
Federal Tax Revenue / GDP	30-37%	

Sources: Instituto de Estudios Fiscales (IEF, 2011). The top marginal rate of the individual income tax was reduced to 46% 2005. The top marginal rate of the corporate income tax was reduced to 32.5% in 2006 and 30% in 2007. The data on tax revenues reflects averages for the period 1999-2007 and include regional-level revenues in all calculations.

Table 8: Bunching Estimation by Year, 2008-2011

	b_{NF}	b_F	B	H	y_{lb}	y_{ub}	N
Pooled data							
2008-2011	.028***	.215***	476.1	556.8	5.50	6.30	122,521
	(.004)	(.043)					
Annual data							
2008	.046***	.225***	240.8	268.5	5.50	6.30	37,393
	(.006)	(.043)					
2009	.017**	.309	75.4	85.6	5.50	6.42	31,430
	(.008)	(.443)					
2010	.014**	.105	58.2	81.9	5.50	6.18	30,191
	(.006)	(.061)					
2011	.028***	.200***	95.1	118.8	5.50	6.30	23,507
	(.009)	(.087)					

Appendix Figures

Figure 12: Behavioral (Non)response at the Corporate Income Tax Threshold

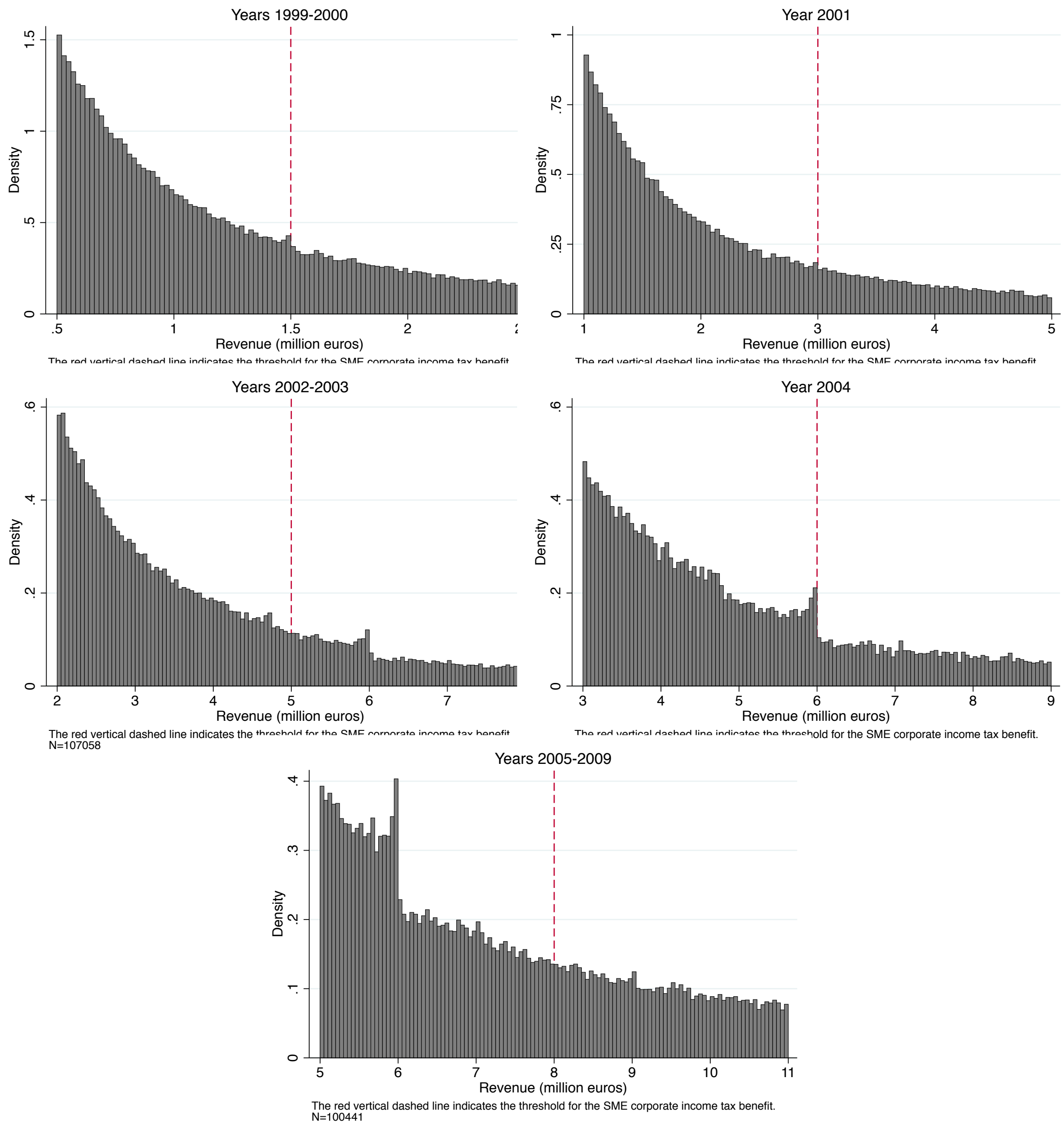
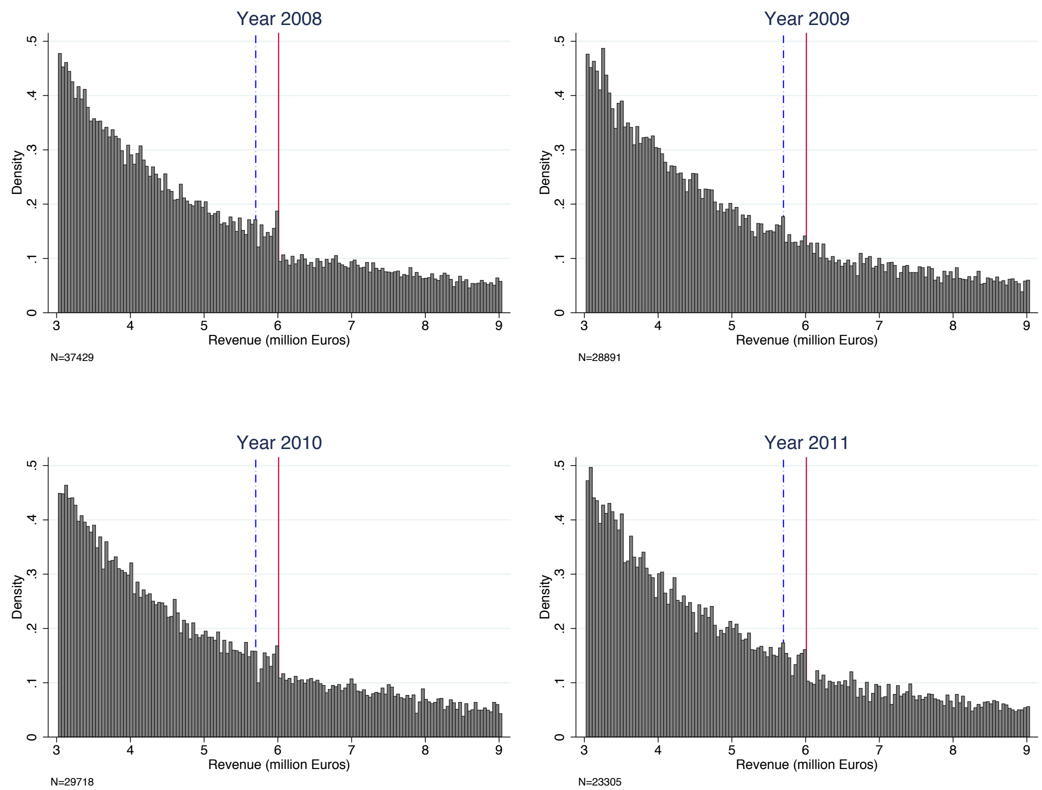
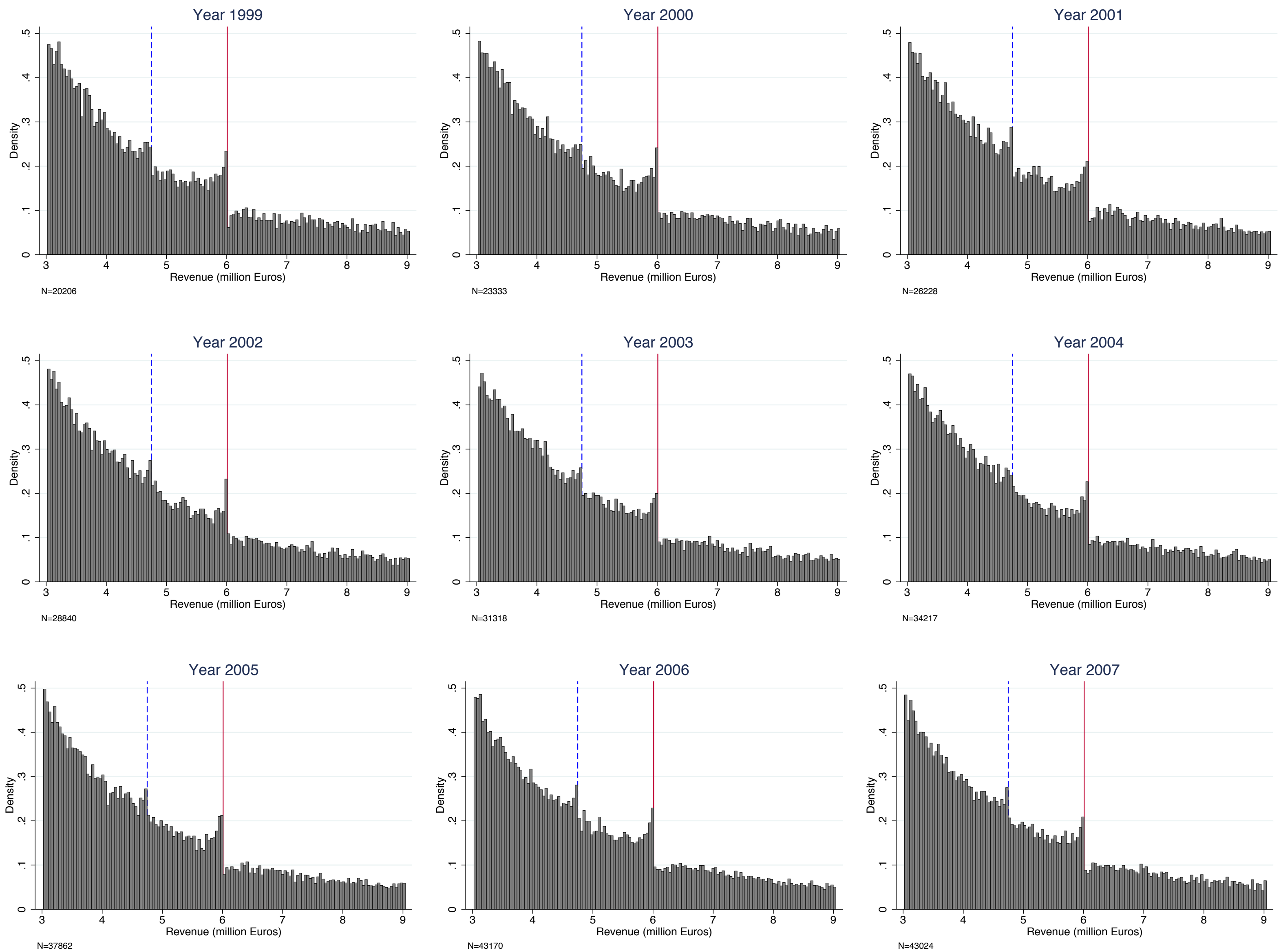


Figure 14: Revenue Distribution, Year by Year, 2008-2011



Note: this figure shows annual histograms of reported revenue for each year in the period 2008-2011. The distribution is very similar for all years, with some noise due to the fact that these subsamples are relatively small. The dashed (blue) line indicates the External Audit threshold, set at €5.7 million during 2008-2011. The solid (red) line indicates the LTU threshold, set at €6 million for both periods (and fixed in nominal terms). The bins are exactly €42,070 wide and are defined such that no bin contains data both to the left and to the right of each threshold.

Figure 13: Revenue Distribution Year by Year, 1999-2007



Note: this figure shows annual histograms of reported revenue for each year in the period 1999-2007. The dashed (blue) line indicates the External Audit threshold, set at €4.75 million during 1999-2007. The solid (red) line indicates the LTU threshold, set at €6 million for both periods (and fixed in nominal terms). The bins are exactly €42,070 wide and are defined such that no bin contains data both to the left and to the right of each threshold.

Figure 15: Operating Revenue Distribution, period 2008-2011

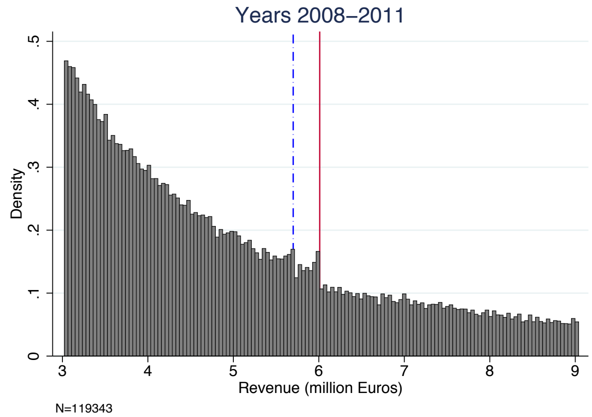


Figure 16: Exporters vs. Non-Exporters

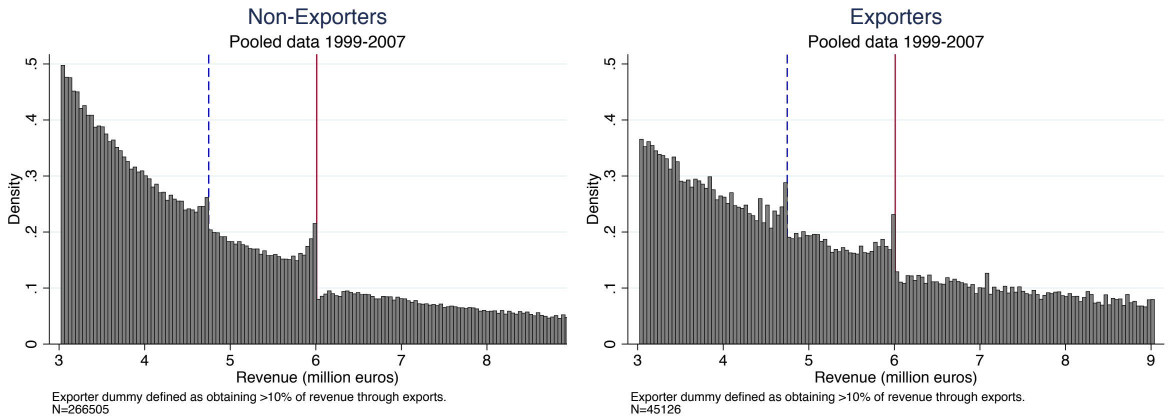
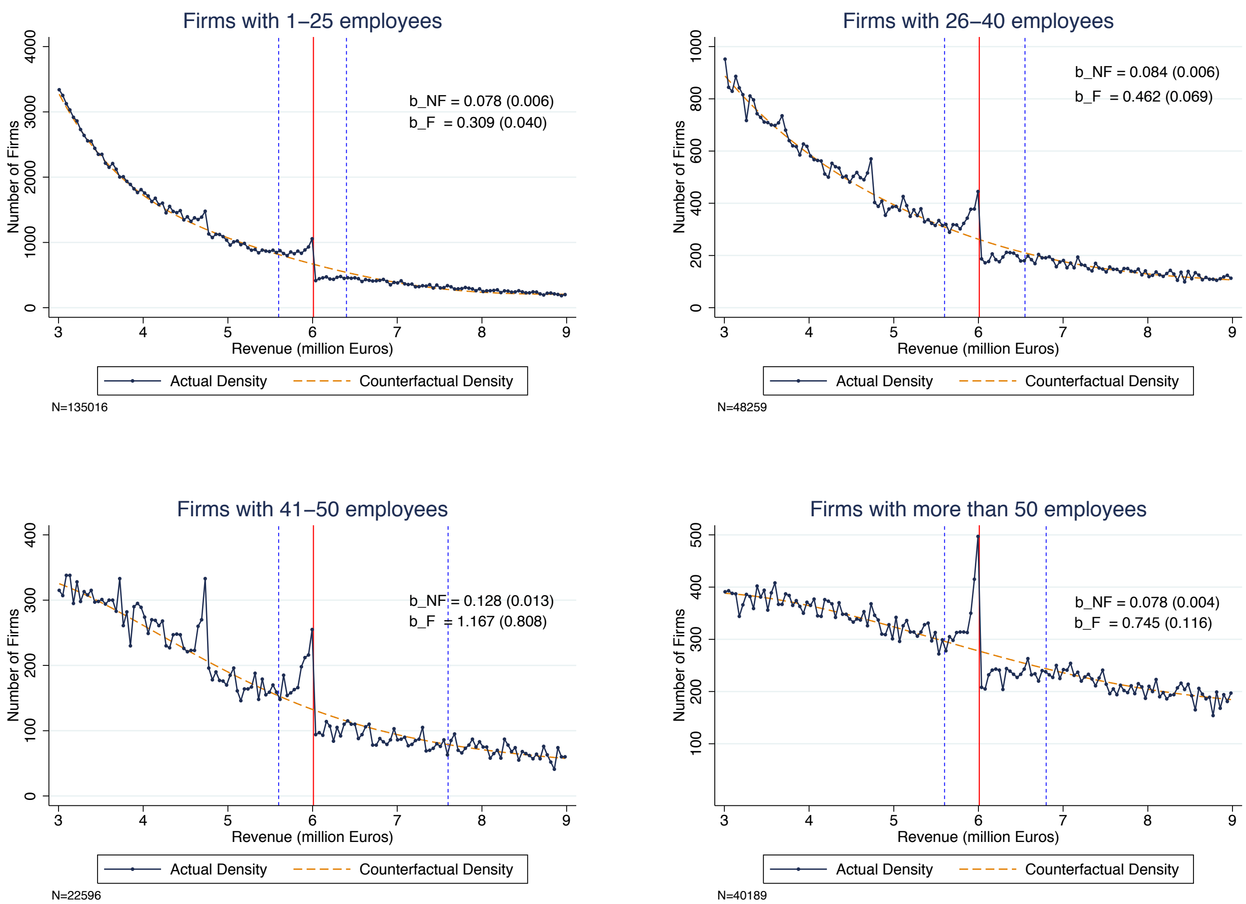
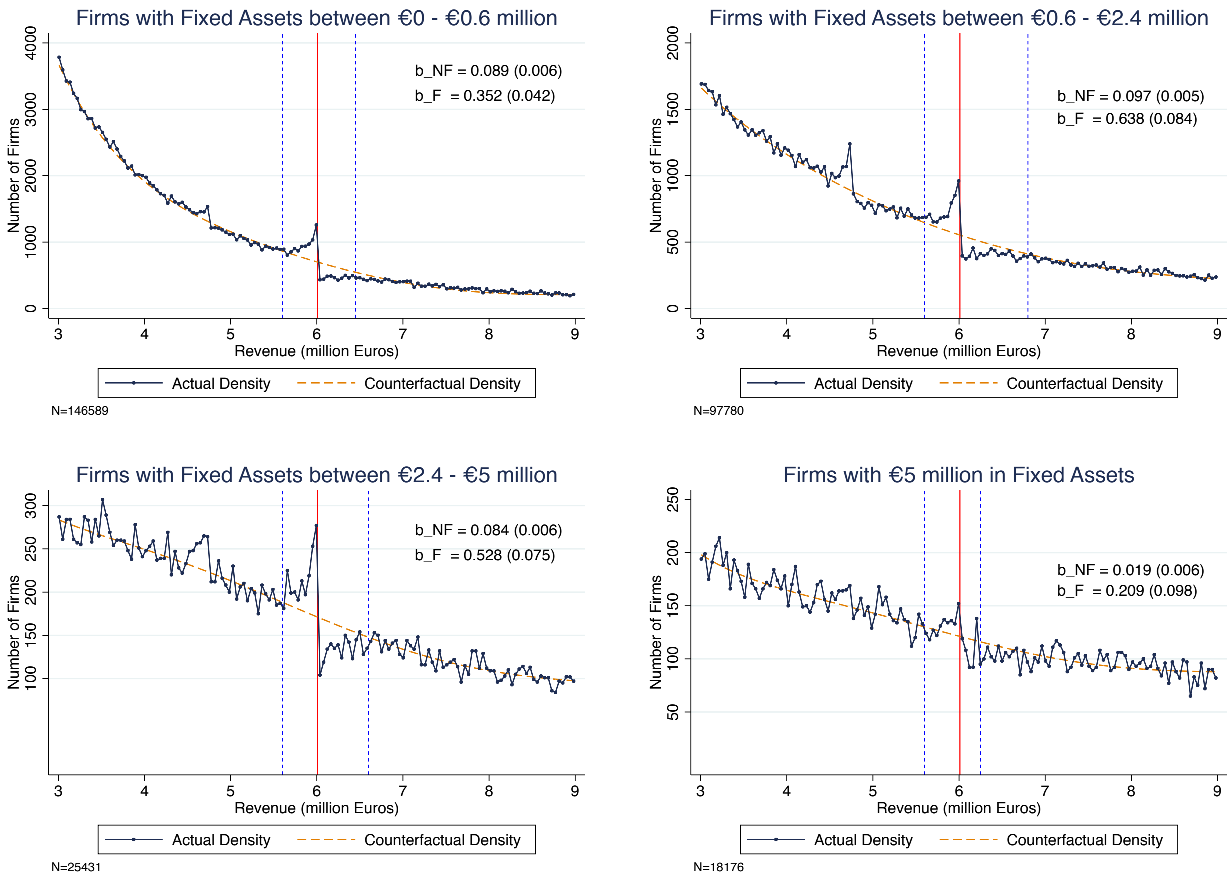


Figure 17: Revenue Distribution by Number of Employees



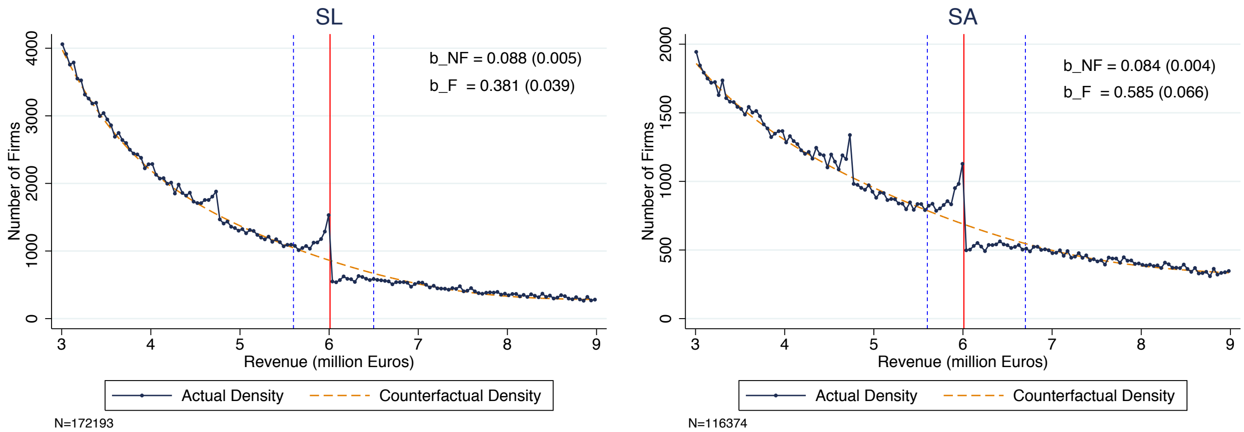
Note: these graphs show the actual and counterfactual revenue distributions for subsamples of firms with a given number of employees. The counterfactual distributions are constructed as explained in the note to Figure 3. Only data for the period 1999-2007 is used.

Figure 18: Revenue Distribution by Fixed Assets



Note: these graphs show the actual and counterfactual revenue distributions for subsamples of firms with a given level of fixed assets (measured in million Euros). The counterfactual distributions are constructed as explained in the note to Figure 3. Only data for the period 1999-2007 is used.

Figure 19: Revenue Distribution by Organizational Form



Note: these graphs show the actual and counterfactual revenue distributions for firms with different organizational forms. SL stands for *Sociedad Limited*, equivalent to a Limited Liability Company. SA stands for *Sociedad Anónima*, equivalent to a Corporation. The counterfactual distribution is constructed in each case as explained in the note to Figure 3. Only data for the period 1999-2007 is used.

Figure 20: Bunching Intensity by Region

Note: this maps represents the 17 Autonomous Regions of Spain. We use a color scale to show the different bunching intensity observed in the revenue distribution in each region, according to the bunching parameter b_{NF}. Lighter (yellow) tones apply to low bunching regions, while darker tones (red) denote high-bunching regions. The lowest bunching is observed in Navarra and País Vasco, the two regions in the North where the Large Taxpayers' Unit (LTU) only applies to a some firms (those that operate extensively in the rest of the country). For the other regions, the pattern is: relatively low bunching in the Northern and Eastern regions (Cataluña, Aragón, Valencia and Baleares) and relatively high bunching in the South, Center and North-West.