Title
Size-Dependent Policies and Firm Behavior

Permalink
https://escholarship.org/uc/item/2b37087g

Author
Almunia Candela, Miguel

Publication Date
2013

Peer reviewed|Thesis/dissertation

Size-Dependent Policies and Firm Behavior

by

Miguel Almunia Candela

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Emmanuel Saez, Chair
Professor Alan Auerbach
Professor Edward Miguel
Professor Ernesto Dal Bó

Spring 2013

Size-Dependent Policies and Firm Behavior

Copyright © 2013

by

Miguel Almunia Candela

Abstract

Size-Dependent Policies and Firm Behavior

by

Miguel Almunia Candela

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Emmanuel Saez, Chair

Most countries have laws that offer regulatory advantages to small firms, such as lower taxes or more flexible labor rules. To determine what firms are eligible to these advantages, it is necessary to define what characterizes a small firm. This is usually done by specifying thresholds in terms of the maximum number of employees, annual revenues, total assets, or a combination of all three. The existence of such thresholds gives firms incentives to strategically remain small to benefit from the regulatory advantages. It also provides researchers with an opportunity to analyze the effects of those regulations, by studying the behavior of firms that are close to the eligibility cutoff.

In the first chapter, my co-author David Lopez-Rodriguez and I study the effects on firm behavior of a discontinuity in tax enforcement intensity in Spain. The Large Taxpayers' Unit (LTU), established in 1995, monitors and enforces the taxation of companies with operating revenue above €6 million, resulting in more frequent tax audits and more information requirements for those firms. We exploit this discontinuity to estimate the impact of tax enforcement on firms' reporting behavior, using a panel dataset of financial statements for Spanish firms from the period 1999-2011. We find an excess mass of firms locating, or "bunching", just below the revenue threshold. Bunching is stronger in the boom period (1999-2007) than in the recession period (2008-2011). Based on the number of bunching firms, we estimate that firms reduce reported revenue by up to 7.5% in the boom period 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, and there is short-term persistence (up to 3 years) in bunching behavior.

In the second chapter, my co-author David Lopez-Rodriguez and I analyze whether bunching of firms below a discontinuity in tax enforcement intensity is due to production (real) or evasion responses. Using an extended theoretical framework, we derive predictions about the behavior of reported input costs under the polar hypothesis of a pure real response and a pure evasion response. We test the plausibility of the two hypotheses using graphical evidence on the patterns of reported input costs around the LTU threshold. This evidence suggests that bunching firms underreport their revenue, overreport their material input costs

and underreport their labor costs in order to evade several taxes: corporate income tax, payroll tax and the value added tax (VAT). We also run panel regressions with firm fixed effects which broadly confirm the results from the graphical analysis. Overall, the results suggest that firms react to this tax enforcement policy mostly through changes in reporting, rather than changes in production. The efficiency costs of tax enforcement are thus likely to be small because tax evasion constitutes a reallocation of income to tax-evading firms, rather than a net loss for society. Finally, we do a rough estimation of the upper bound of corporate income tax evasion, which yields a modest amount of evasion.

In the third chapter, I study the impact of a set of labor regulations in France that applies only to firms with more than 50 employees. These regulations increase the average labor cost per employee, giving firms an incentive to remain small. The firm size distribution shows strong bunching below the threshold for the period 2002-2008. In terms of growth patterns, the proportion of firms increasing their size from one year to the next drops almost by half at 49 employees, while the share of firms keeping employment constant doubles. I set up a stylized model where firms only choose their number of employees to derive an expression for the elasticity of labor demand, and then estimate it using the number of bunching firms as a sufficient statistic. I obtain a point estimate of $e = 0.055$, which is statistically different from zero at the 10% level. Making an adjustment for the possibility that some firms do not respond to the regulations due to optimization frictions, I obtain a point estimate of $e_F = 0.572$. The latter can be interpreted as an upper bound for the long-term structural elasticity, although it is imprecisely estimated (the standard error is 0.668). These point estimates are considerably below labor demand elasticities estimated in the literature, which according to Hamermesh (1993) tend to be in the interval $(0.15, 0.75)$. An intuitive explanation for why I obtain low point estimates is that bunching firms may be adjusting their production by increasing the use of other inputs instead of labor. I find some preliminary evidence supporting this hypothesis: median fixed assets per employee drop sharply at the notch, indicating that bunching firms have a higher capital-labor ratio than firms just above the threshold.

To my parents, Mila and Joaquín, for inspiring my passion for learning;

and to Paola, for inspiring me to live, love, and smile every single day.

Contents

List of Figures

List of Tables

Acknowledgements

his good spirits in and out of the tennis courts, Gianmarco and his genuine surfer vibe... Last but not least, I was also very lucky with my officemates: Ana, Alex, Attila, Michael and Owen. Thank you for all the interesting discussions and the fun days at work, despite the lack of windows and the disgusting carpet.

My great friend and co-author David Lopez-Rodriguez deserves special recognition. Who would have thought that a casual conversation about the Spanish economic crisis would lead to such a long (and not yet over!) research journey? I am extremely grateful for all the hard work and all the Skype hours that he has put into our joint research project, even though he knew he could free-ride because I had strong incentives to do all the work myself. I know few people as generous and hard-working as David. Gracias, compañero!

I would not be writing these lines today if it weren't for my parents, Mila and Joaquín. (Well, obviously. I mean beyond purely biological reasons.) Together with my older sister, Cristina, they taught me to think for myself and to never take conventional wisdom for granted. They also taught me to believe in my own potential and to be ambitious. And they made sure I learned English as a teenager, even though I'd rather not leave my small neighborhood at the time. As if all of that wasn't enough, in the past few years they followed every step of the process, asking how my exams went, whether I was making progress with my research, and celebrating my little successes along the way as if they were their own.

Last but not least, a special and infinite thank you goes to the love of my life, Paola. Our paths crossed in a very unexpected place and time, and she has made the last four years the happiest period of my life. She supported me in so many ways, taking care of me when I had a lot of work to do, listening to my ideas and my frustrations, helping me stay optimistic and focused and, most importantly, giving me unconditional love every single day. This dissertation is partly hers.

To all of you: Muchas gracias y hasta pronto!

Berkeley, May 2013

Chapter 1

Firm Responses to Tax Enforcement Strategies: Evidence from a Panel of Spanish Firms

with David Lopez-Rodriguez

1.1 Introduction

Firms remit more than three-quarters of the tax revenues raised by governments in advanced economies.[1] As taxpayers, they remit corporate income tax and their share of payroll tax. As tax collectors, they withhold income and payroll taxes from employees, and in many countries they also remit value added tax (VAT). Despite playing such a crucial role as fiscal intermediaries, the empirical literature on tax evasion has largely neglected firms, focusing instead on individual behavior.[2] 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.

In this paper, we take 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, with their tax schedules remaining unaffected. This discontinuity in tax enforcement intensity gives firms an incentive to remain below the revenue threshold. They can do this either by reducing their output or by underreporting their revenue (or both). In this paper, we estimate the size of the total reported revenue response

---

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

[2]In one of the most authoritative surveys on the issue of tax evasion, Andreoni, Erard and Feinstein (1998) state in the introduction: "[...] Nor do we have the space to discuss corporate or business tax evasion". 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).

without focusing on the distinction between a real production response and misreporting, which is the subject of the second chapter of this dissertation.

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 an exogenous productivity draw that determines their optimal size in equilibrium. The probability of evasion detection 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 for all inframarginal units evaded when a firm crosses the threshold. The existence of the notch drives some firms to report lower revenue and bunch at the LTU threshold to avoid high enforcement. We define the "marginal buncher" as the firm with the highest exogenous productivity that chooses to bunch at the cutoff point.

For 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 85% of registered businesses in Spain with operating revenue above €2 million for the period 1999-2011. 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 an overall picture of the firms' activities, allowing us to observe several dimensions of firm behavior and the impact of multiple taxes in a single data source.[4]

To estimate the firms' response to a discontinuity in tax enforcement intensity, we analyze the distribution of reported operating revenue. 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.

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 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. Several reasons could explain this behavior, for example prior exposure to the LTU or the inability to misreport revenue due to the types of clients served (e.g., government contracts). Another factor could be heterogeneous preferences concerning evasion, as there might be honest business managers who would not evade taxes under any enforcement level. 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.

---

[4]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 governments provide anonymized data).

For the boom period 1999-2007, we find that the marginal buncher reduces its reported revenue by €86,000 (1.4% of total revenue) under the assumption of no optimization frictions, and €449,000 (7.5% of total 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 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.

Even though our bunching estimates are static, we also analyze the dynamic behavior of firms taking advantage of the longitudinal dataset. Growing firms, defined as 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. Moreover, both the probability of revenue growth and the median growth rate drop significantly as firms approach the threshold from below. Finally, we find that 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. Bunching persistence remains statistically significant when the period is extended up to six years, but the economic significance decreases sharply beyond three years.

The empirical techniques used in this paper draw on a recent literature in public finance that analyzes responses to thresholds in taxes and regulations. In his seminal paper, Saez (2010) exploits kink points in the US personal income tax schedule – i.e., income thresholds at which the marginal tax rate jumps – to identify taxable income elasticities.[5] Our empirical strategy draws particularly on the work by Kleven and Waseem (2013), who adapt Saez's method to the case of notches – income thresholds at which the average tax rate jumps. Our setting has two novel characteristics within this literature. First, the LTU generates a notch in the probability of evasion detection, rather than the tax rate (which is unaffected in this setting), which allows us to study tax enforcement in isolation. Second, the notch is defined in terms of operating revenue, which is not a measure of taxable income. The latter adds an extra step in the empirical estimation, because it requires a separate estimation of the effects on revenue and input costs, as explained above.

Finally, this paper contributes to the extensive literature on firm size distribution[6] and size-contingent policies.[7] This topic has received a lot of attention in Spain because of policy reports (e.g., LaCaixa, 2012) showing that Spanish and German firms are equally productive after controlling for firm size (measured by number of employees). The implication is that

---

[5]Several recent studies (Chetty et al. (2011), Chetty, Friedman and Saez (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.

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

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

the entire productivity gap between the two countries is due to differences in the firm size distribution. The findings in this paper suggest that the observed firm size distribution in Spain could be substantially distorted by evasion behavior, raising questions about such productivity calculations.

The rest of the paper is organized as follows. Section 2 presents the theoretical framework. Section 3 provides a description of tax enforcement policies in Spain and of the Amadeus dataset. Sections 4 presents the static empirical analysis, including a derivation of the bunching parameters. Section 5 presents the dynamic empirical analysis. Section 6 concludes.

1.2   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.

1.2.1   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 $\Pi = (1 - t) [\psi f(m) - cm]$. Assuming that tax evasion is not possible, profit maximization yields the standard condition:

$$\psi f_m \left(m^{NoEv} \right) = c \tag{1.1}$$

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 $\overline{\Pi} = (1 - t) [\overline{y} - cm]$. 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:

$$\begin{aligned} \mathbb{E}\Pi &= (1-\delta)\,\Pi^{ND} + \delta\Pi^D \\ &= (1-t)\,\Pi + tu\,[1 - \delta\,(1+\theta)]. \end{aligned} \tag{1.2}$$

1.2.2 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.[8] 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.[9]

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:

$$\psi f_m(m^*) = c + u\left[\frac{t}{1-t}\right][1+\theta]\,\delta_m(u,m^*) \tag{1.3}$$

$$1 = [1+\theta]\,[\delta(u^*,m) + u^* \cdot \delta_u(u^*,m)] \tag{1.4}$$

Condition (1.3) is similar to the standard optimality condition (1.1), 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 (1.3) reduces to (1.1). Comparative statics are intuitive: optimal input use m^* is larger when (i) its effect on the detection probability is weaker (i.e., $\delta_m(u,m)$

---

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

[9]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.

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 (1.4) 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.[10]

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 $\left[\overline{y}^* \left(\underline{\psi} \right), \overline{y}^* \left(\overline{\psi} \right) \right]$.[11]

1.2.3 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.[12] 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

---

[10]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:

$$\frac{du}{d\delta} \big|_{u=u^*} = - \frac{dF/d\delta}{dF/du} \big|_{u=u^*}$$

$$= - \frac{1 + \theta}{[1 + \theta] [\delta_u + \delta_u + u^* \delta_{uu}]} \big|_{u=u^*}$$

$$= - \frac{1}{2\delta_u + u^* \delta_{uu}} \big|_{u=u^*}$$

$$< 0, \text{ since } \delta_u, \delta_{uu} > 0.$$

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

[12]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.

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 (1.3) and (1.4) 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 (1.3) and (1.4). 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.

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 $f\left(m\right)$ and $\delta\left(u,m\right)$. 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 $\left(\delta^{LTU}\right)$:

$$\mathbb{E}\Pi\left(u^{**}, m^{**}|\psi^{MB}, \delta\right) = \mathbb{E}\Pi\left(u^{**}, m^{**}|\psi^{MB}, \delta^{LTU}\right) \qquad (1.5)$$

An important point to notice about expression (1.5) 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 (1.3) and (1.4), but with $\delta^{LTU}\left(u,m\right)$ instead of $\delta\left(u,m\right)$. Hence, these "high productivity" firms re-optimize and report higher revenue than they did in the benchmark case: $\overline{y}^{**}\left(\psi > \psi^{MB}\right) > \overline{y}^*\left(\psi > \psi^{MB}\right) > y^{LTU}$.

7

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 $d\delta \equiv \delta^{LTU}\left(\cdot\right) - \delta\left(\cdot\right)$. 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, $h_0\left(\psi\right)$, to the pdf of reported revenue, $g_0\left(\overline{y}\right)$. Hence, we can define the excess mass of bunching firms, B, as follows:

$$B = \int_{y^{LTU}}^{y^{LTU}+d\overline{y}} g_0\left(\overline{y}\right) d\overline{y} \approx g_0\left(y^{LTU}\right) d\overline{y}^{MB}, \tag{1.6}$$

where the approximation assumes that the counterfactual density $g_0\left(\overline{y}\right)$ 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.[13] Under the strong assumption that firms face no optimization frictions[14], $d\overline{y}^{MB}$ 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 (1.6) to define the parameter b as the ratio of excess bunching over the counterfactual density at the threshold:

$$b \equiv \frac{B}{g_0\left(y^{LTU}\right)} \approx d\overline{y}^{MB} \tag{1.7}$$

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.

1.3   Institutional Context and Data

Tax agencies around the world monitor large taxpayers more closely than small ones. This policy is justified because the expected tax revenue recovered is higher when monitoring

---

[13]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 $\overline{y}^{**} = y^{LTU}$.

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

large taxpayers, even considering that expected enforcement costs per taxpayer (e.g., the cost of conducting tax audits, requesting and processing information) increase with firm size. Most OECD countries have some type of Large Taxpayers' Unit (LTU) dedicated exclusively to monitoring and enforcing taxes on the largest companies (OECD, 2011). International institutions like the IMF have supported the establishment of LTUs in developing countries over the last 20 years, arguing that they improve enforcement policies and increase tax revenue (IMF, 2002). By analyzing the impact of the Spanish LTU on firm behavior, we provide some new evidence that may be applicable to other contexts.

We summarize below the key characteristics of the Spanish LTU, the main source of variation exploited in the empirical section of this paper. We also describe a second policy threshold above which firms are required to hire a external private firm to audit their annual accounts. Finally, we describe the data used in the empirical sections of the paper.

1.3.1 Tax Administration Thresholds in Spain

1.3.1.1 LTU threshold

The Spanish tax agency established a LTU (*"Unidades de Gestión de Grandes Empresas"*) in 1995 to closely monitor tax compliance by the largest firms operating in the country. The threshold to define a "large firm" was set at €6 million[15] in annual operating revenue and has not been modified since then.[16] When a firm reports revenue above the threshold in a given year, it is automatically added to a 'census' of large firms starting the *following* year. Exporters are always included in the LTU, regardless of their total revenue, because they can potentially claim large VAT reimbursements on their exports.

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.[17] In terms of compliance requirements, firms in the LTU census are required to file their value-added tax declarations on a monthly basis (instead of quarterly) and in electronic form (as opposed to on paper).[18] Moreover, the withholding rate on the corporate income tax is 25%, compared to 18% for small firms.[19] To summarize, (i) firms in the LTU are more likely to be audited, (ii) it is easier to cross-check their individual transactions, and (iii) they may face liquidity constraints due to more frequent and higher tax withholding.

Over time, the composition of the LTU Census has changed. While the threshold has remained fixed in nominal terms, inflation (approximately 3% per year) has brought many

---

[15]The threshold was originally set at 1 billion pesetas, the official currency at the time. The official exchange rate is 166.386 pesetas per euro, so the threshold is exactly €6.010121 million (no rounding was applied). All the graphical evidence below specifies the exact threshold.

[16]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 large firms within the LTU. We observe no evidence of bunching at this revenue threshold, partly because the mass of firms around this revenue level is quite low.

[17]As reported by AEAT (the Spanish tax agency) in its Annual Reports AEAT (Several years).

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

[19]To be precise, the withholding rate for firms in the LTU firms is 5/7 of the statutory rate, yielding $35 * \frac{5}{7} = 25\%$ for most firms. For companies below the threshold, the withholding rate is exactly 18% (BOE, Several years). Post-2007 reforms have modified these rates.

firms above the cutoff, even if they were not growing in real terms. Combined with a 3% average annual GDP growth rate, the number of companies in the LTU census increased from 18,860 (2.4% of all registered firms) in 1999 to 40,571 (2.9%) in 2007. Firms in the LTU report more than 80% of all business profits and about 75% of their share of *taxable* profits is around (AEAT, Several years). The magnitude of those number is due to the fact that firms in the LTU are the largest and most productive in the economy, while the discrepancy suggest that these firms take advantage of more tax deductions. In the period under study, overall LTU staff stayed essentially constant, but there were substantial technical improvements, so the net change in enforcement intensity over time is likely to be limited.

The LTU has one office in each of the 17 Spanish autonomous regions (*Comunidades Autónomas*). The two largest offices are located in the regions of Madrid and Cataluña, where the two largest cities (Madrid and Barcelona) are located. Each regional LTU team is only responsible for monitoring the firms whose headquarters are located in the region. Teams are given annual targets in terms of total firms audited, but we do not have data to exploit potential variation in the effectiveness of each regional team.

1.3.1.2 Other Tax Administration Thresholds

There are two other thresholds that could affect the behavior of firms around the LTU threshold. We first describe the External Audit threshold and then a Corporate Income Tax 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;[20] 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. These criteria were modified starting in 2008, raising the revenue threshold to €5.7 million and the assets threshold to €2.85 million. Despite not being implemented directly by the tax agency, this size-dependent requirement complements tax enforcement policy because official tax audits typically use the private auditor's report as a source of information. Private auditors are required by law to provide a "truthful assessment of the company's accounting", and they face legal responsibility if any misreporting is found. For this reason, auditors are wary to sign an audit report if they find obvious evidence of tax evasion, which limits the ability to evade for audited firms.[21] The fee charged by private auditors varies with the size of the business and the complexity of its operations. For a firm with revenue close to €4.75 million, the average costs during the period under study was in the range €10,000 - €30,000, a small but non-negligible expenditure (0.2 to 0.6% of total revenue, but 4 to 12% of reported *profits* on average). Beyond the monetary fee, a private audit implies administrative costs to the firm related to compiling information and dealing with the external auditor.

The Corporate Income Tax threshold offers a marginal tax rate of 30% (instead of the

---

[20]As with the LTU threshold, these amounts were established before the adoption of the Euro. The revenue limit was originally 790 million Pesetas (€4.748 million), and the assets limit was 395 million Pesetas (€2.374 million).

[21]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.

standard 35% rate) to qualifying firms.[22] Eligibility for this tax break is determined exclusively by a revenue criterion, like the LTU. However, the threshold for the tax break was modified every few years to account for nominal economic growth.[23] The eligibility cutoff changed over time from €1.5 million in 1999 up to €10 million in 2010 (full details are provided in Table 1.2). The cutoff for the tax break overlapped with the LTU threshold during 2004, but was different in all other years.

1.3.2 Data

We use firm-level data from Amadeus, a comprehensive database of European businesses put together by Bureau van Dijk, a market research company (www.bvdinfo.com). The dataset covers annual accounting reports for the period 1999-2011.[24] All firms in Spain are required by law to deposit their annual accounts at the Commercial Registry (*Registro Mercantil Central*). Amadeus compiles all these annual reports into a longitudinal database. The information available for each firm in each year includes: business 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. Some of the key variables that we use in the empirical section are: net revenue from sales, (end-of-year) number of employees, total wage bill, and total expenditures on material inputs.

The main advantage of this dataset is the panel structure, which allows us to study the behavior of firms over time, facing the same policy thresholds repeatedly. Another important aspect is that it allows us to 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 not the case when researchers obtain access to administrative data, because these are often anonymized data that cannot be linked to other source.

The dataset also has some limitations. First, a large number of small firms do not fulfill the reporting requirement because it is costly to them and the associated fines are small. The main advantage of complying is that submitting the annual accounts is a usual requirement to obtain loans from commercial banks and government contracts. We compare the size of the Amadeus dataset to the number of firms submitting corporate income tax returns to the tax agency. Amadeus contains information from approximately 85% of firms with annual revenue between €1.5-€60 million that submitted a corporate tax return to the Spanish tax agency. The percentage is close to 90% for firms larger than €60 million, but just below 50% for firms smaller than €1.5 million. Table 1.1 shows the comparison between the two data sources. This study focuses on firms with revenue between €2-€12 million, so we treat the available data as the quasi-universe of Spanish firms in that size range, which corresponds to a small-medium enterprise size. Assuming that missing firms are more likely to be tax

---

[22]The lower rate applied only up to the first €90,121 of profits (this cutoff was modified later, as shown in Table 1.2). For profits above that level, the marginal tax rate is 35% even for firms eligible to the tax break. This creates a kink in the corporate income tax for qualifying firms. Restricting the sample to these firms, we observe a small bunching response around this kink.

[23]There were also political motivations behind these reforms, because extending tax breaks for small businesses is usually a popular policy.

[24]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 changes.

evaders than those included in Amadeus, the worst possible scenario is that selection bias would make our estimations of tax evasion a lower bound.

A second limitation, common to the corporate tax literature, is that the financial statements may 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.[25] To know the exact tax liability, we would need administrative tax return data for all the major taxes, 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. With these caveats in mind, we consider these data to be almost as good as administrative data for the small and medium-size firms we study in the following sections.

1.4 Empirical Strategy: Static Analysis

1.4.1 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 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 bunching of firms at the revenue thresholds described above indicates a behavioral response to tax administration policies. We separately analyze two periods: 1999-2007, when the economy was booming, and 2008-2011, when the economy was in recession.[26]

Using data from Amadeus, the top panel of Figure 1.2 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 smaller spike at the External Audit threshold suggests that complying with this administrative requirement is less costly to firms than facing higher tax enforcement. However, since the criteria to determine eligibility for the external audit involve other variables apart from revenue, it is difficult to draw strong conclusions. For this reason, in the remainder of this paper we focus almost exclusively on

---

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

[26] Real annual GDP growth was 3.5% in the period 1999-2007, while inflation remained above 3% per year. Hence, nominal annual growth was close to 7% in that period. In contrast, growth was on average -0.5% in the period 2008-2011, especially due to the sharp fall of -3.7% in 2009. Inflation was around 2% per year, so annual nominal growth was only 1.5% in the second period.

the response to the LTU threshold.

The pattern for the recession period 2008-2011 is quite different. 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.

There are essentially two ways in which firms can reduce their reported revenues in order to bunch at one of the thresholds. They can produce less output or charge lower prices, which we refer to as "real" responses because they both reduce the true revenue raised by the company. Alternatively, firms can underreport their sales, either my misreporting the amount sold or the unit price. We refer to the latter as "evasion" responses, because they reduce the firm's tax liabilities both on the corporate income tax and the VAT.

Regardless of the type of response, firms incur a cost in order to strategically locate below the threshold. In the case of real responses, the cost is related to moving away from the optimal production level that maximizes true after-tax profits. This is a pure efficiency cost, because it puts total output below the social optimum. In the case of evasion responses, firms may face costs like foregoing business opportunities due to operating in cash, or the additional costs of keeping two sets of accounting books (one for internal use and one to show the tax agency). These are examples of resource costs of evasion, which also involve a loss of efficiency because firms spend resources in an unproductive way. As pointed out by Chetty (2009a), there are also transfer costs of evasion, such as monetary penalties paid if detected. The penalty represents a private cost to the taxpayer, but has no social cost because it is a transfer to the government. If taxpayers face only transfer costs but no resource costs of evasion, there would be no efficiency costs related to evasion responses.

Disentangling real and evasion responses empirically is challenging in this setting, because they are observationally equivalent in terms of the revenue distribution. The remainder of this paper will focus exclusively on quantifying the total response of reported revenue. This analysis allows us to put an upper bound on each type of response. Chapter 2 of this dissertation discusses possible empirical strategies to assess the importance of the two potential responses and the consequences for welfare.

There are two relevant concerns about the interpretation of the graphical evidence obtained pooling multiple annual cross-sections. First, there may be heterogeneity in the bunching response across years that could get hidden in the aggregate picture. Figures 1.3 and 1.4 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. In subsection 4.2, we analyze why these two periods lead to different patterns of bunching behavior.

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 in section 3.1.2. The annual revenue distributions plotted in Figures 1.3 and 1.4 show no discernible bunching at the tax break threshold in any year other than 2004 (the only year when the two threshold coincide). Hence, we conclude that firms do not respond to this tax incentive in a significant way. The lack of reaction to a 5-percentage-point (14 *percent*) 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.

1.4.2 Quantifying the Bunching Response

In order to quantify the size of the response to the LTU threshold, 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[27] and estimate the following polynomial regression:

$$F_j = \sum_{i=0}^{q} \beta_i \cdot (y_j)^i + \sum_{k=y_{lb}}^{y_{ub}} \gamma_k \cdot \mathbb{1}\,(y_j = k) + \eta_j \tag{1.8}$$

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.[28]

We estimate the counterfactual distribution by calculating predicted values with the estimated coefficients from regression (1.8), excluding the γ_k shifters to eliminate the perturbations around the threshold. Hence, the counterfactual density is given by:

$$\widehat{F_j} = \sum_{i=0}^{q} \widehat{\beta_i} \cdot (y_j)^i \tag{1.9}$$

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

---

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

[28]In this particular application, we add to equation (1.8) 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.

right of the threshold (H):[29]

$$\widehat{B} = \sum_{j=y_{lb}}^{y^{LTU}} \left| F_j - \widehat{F_j} \right| \qquad \widehat{H} = \sum_{j=y^{LTU}}^{y_{ub}} \left| F_j - \widehat{F_j} \right| \qquad (1.10)$$

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 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 (1.8) 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 (1.8) again as long as $\widehat{B} > \widehat{H}$. The process continues until it reaches a value of y_{ub} such that $\widehat{B} = \widehat{H}$.[30]

The results obtained allow us to estimate the bunching parameter b defined in equation (1.7), 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:

$$\widehat{b}_{NF} = \frac{\widehat{B}}{\left[\frac{1}{1+(y^{LTU}-y_{lb})/w}\right] \sum_{j=y_{lb}}^{y^{LTU}} \widehat{\beta}_i \cdot (y_j)^i}, \qquad (1.11)$$

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 1.2 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 (1.8) a large number of times (with replacement) to obtain bootstrapped standard errors.[31]

---

[29]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.

[30]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: $\left(\widehat{H}/\widehat{B}\right) > 0.95$.

[31]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.

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 statistically 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 (1.9) 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.[32] 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 1.4 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.

1.4.2.1 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.

---

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

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 1.6 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.[33] We conclude that some growing firms avoid the

---

[33]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.

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.[34] 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:

$$\widehat{b}_F = \frac{\widehat{b}_{NF}}{1 - \alpha} \tag{1.12}$$

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. This issue is tackled in chapter 2 of this dissertation.

1.4.3   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 1.5 and Figures 1.7-1.12.

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.

---

[34]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.

Figures 1.7 and 1.8 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.[35] 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.

Table 1.5 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 1.9. 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 1.10 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

---

[35]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.

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). We define a "scope of evasion" index at the sector level with two components. First, we obtain the percentage of a sector's output that is sold to final consumers from the input-output tables of the Spanish economy.[36] Second, we calculate the median number of employees for each sector, using only firms with revenue between €5.5-€6.5 million. The intuition for the first element 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. Formally, we compute:

$$Scope = ShareFinalCons * (1 - employees/100)$$

where we normalize the number of employees so that both terms are smaller than one and $Scope \in (0,1)$.

Figure 1.12 shows a scatterplot of bunching responses (measured by $\widehat{b}_{NF}$, vertical axis) and the scope of evasion index (horizontal axis), to illustrate their relationship for different sectors in the economy. There is not an obvious pattern, but it seems clear that a linear fit is not the best way to approximate the relationship. Instead, we have overlaid a quadratic fit that shows an inverted-U pattern. One way to interpret the relationship is to consider the case of extreme data points. In the sector with lowest scope of evasion (restaurants and hotels), the bunching response is very small. This might be explained by the fact that these are "large" firms in terms of employees, which makes misreporting more difficult. On the other extreme are retailers, which have the highest scope for evasion but do not bunch very strongly. One hypothesis could be that when retailers underreport, it is extremely hard even for the LTU to detect it because 85% of sales go to final consumers (and most of them are untraceable 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.

1.5 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 1.13 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%

---

[36]The input-output tables are published by the National Statistics Institute.

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.[37] The top panel of Figure 1.14 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 1.14.

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.[38] 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.[39] 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:

$$Prob\left[bin\left(y_{it}\right) = bin\left(y_{i,t+h}\right)\right] = \alpha + \beta BunchBin_{it} + y_{it} + y_{it}^2 + \lambda_t + \varepsilon_{it} \qquad (1.13)$$

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

---

[37]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.

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

[39]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 1.15 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 $y_{it} \in (\text{€}2, \text{€}2.25$ million$)$ to 6% in the range $y_{it} \in (\text{€}9.75, \text{€}10$ 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 1.6. 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).

1.6   Concluding Remarks

This paper has documented the response of Spanish firms to a discontinuity in tax enforcement intensity created by the Large Taxpayers' Unit (LTU), which monitors and enforces the taxation of firms with revenue above €6 million. Throughout the period 1999-2011, the distribution of reported revenue features bunching below the LTU threshold. Static bunching estimates similar to those developed in the earlier literature indicate that bunching firms strategically reduced their reported revenue by between €86,000 (1.4%) and €446,000 (7.5%) in the boom period 1999-2007. The estimates are significantly smaller (€26,000 and €384,000), but still statistically significant, for the recession period 2008-2011.

A dynamic analysis of firm behavior shows that almost all the bunching is due to growing firms, defined as businesses whose reported revenue increased from the previous to the current year. This partly explains why bunching behavior changes so clearly after the economic cycle changes. We document how growth patterns change as firms approach the LTU threshold from below, and also show that the degree of bunching persistence is substantial when looking at one to three-year horizons, but fades away when looking at longer time periods. This reflects the high costs of bunching when the nominal threshold loses value in real terms due to inflation (which averaged 3% in Spain throughout the period under study).

We have abstracted from evaluating whether having a clear-cut revenue threshold to delimit two enforcement regimes is an optimal policy from the mechanism design point of view. An alternative policy would be to expand the LTU by lowering the threshold so that more firms fall under its jurisdiction. This would have high short-term returns, because the LTU would now target bunching firms, which are the most likely to be evading taxes. However, over time firms are likely to learn that the threshold has moved and would adjust their strategic behavior to bunch at the new threshold. Hence, it seems like a more effective reform would be to make the threshold "fuzzy". That is, make the transition to the high enforcement regime happen over a relatively large revenue interval, rather than a fixed threshold. A smooth increase in the perceived audit probability would suffice to eliminate the incentives to bunch at a given revenue level. In practical terms, this could apply mostly to audit rates, but not to other administrative LTU elements such as extra compliance

requirements. With a fuzzy threshold, the firms' decision would be more uncertain and, at the same time, less salient because there would be no point at which the perceived probability of detection changes sharply.

Tables

Table 1.1: 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 1.2: Revenue Threshold for Corporate Income Tax Break for Small Firms

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

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 1.3: Bunching Estimations

	b_{NF}	b_F	B	H	y_{lb}	y_{ub}	N
Pooled data							
1999-2007	.086***	.465***	3553.1	3576.6	5.50	6.60	291,217
	(.005)	(.052)					
2008-2011	.026***	.384***	442.9	466.9	5.50	6.25	121,277
	(.004)	(.042)					
Annual data							
1999	.142***	.745***	395.1	377.7	5.50	6.95	20,417
	(.015)	(.177)					
2000	.117***	.567***	386.3	378.8	5.50	6.75	23,534
	(.012)	(.128)					
2001	.085***	.520***	316.5	336.8	5.50	6.75	26,471
	(.011)	(.133)					
2002	.060***	.388***	249.9	259.0	5.50	6.50	29,095
	(.009)	(.095)					
2003	.071***	.334***	319.0	327.4	5.50	6.45	31,646
	(.009)	(.071)					
2004	.090***	.438***	438.5	436.0	5.50	6.55	34,565
	(.008)	(.069)					
2005	.076***	.384***	413.8	439.5	5.50	6.50	38,229
	(.008)	(.065)					
2006	.090***	.476***	562.3	559.1	5.50	6.60	43,718
	(.008)	(.083)					
2007	.076***	.425***	474.6	457.8	5.50	6.50	43,542
	(.008)	(.076)					
2008	.040***	.184***	213.1	251.1	5.50	6.25	37,391
	(.006)	(.038)					
2009	.014*	.263	57.7	77.1	5.50	6.45	29,262
	(.009)	(.598)					
2010	.009	.069	39.6	123.7	5.50	6.25	30,326
	(.008)	(.073)					
2011	.031***	.217***	105.5	121.9	5.50	6.30	23,758
	(.009)	(.095)					

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 1.4: 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	6.55	0.081***	0.425***
5.40	6.60	0.084***	0.453***
5.50	6.60	0.086***	0.465***
5.60	6.55	0.086***	0.452***
5.70	6.50	0.083***	0.425***

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 1.5: 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***
$\left(\overline{y}_t > \overline{y}_{t-1}\right)$	(.004)	(.040)
Shrinking	.015	.940
$\left(\overline{y}_t < \overline{y}_{t-1}\right)$	(.014)	(1.674)

Note: this table reports the bunching intensity estimates for "no frictions" and "frictions" $\left(\widehat{b}_{NF}\right.$ 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 1.6: Bunching Persistence Over Time: Regression Results

	$h{=}1$	$h{=}2$	$h{=}3$	$h{=}4$	$h{=}5$	$h{=}6$	$h{=}7$	$h{=}8$
	Dependent Variable: $Pr\left[bin\left(y_t\right)=bin\left(y_{t+h}\right)\right]$							
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:

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

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 1.1: Theoretical Revenue Distribution
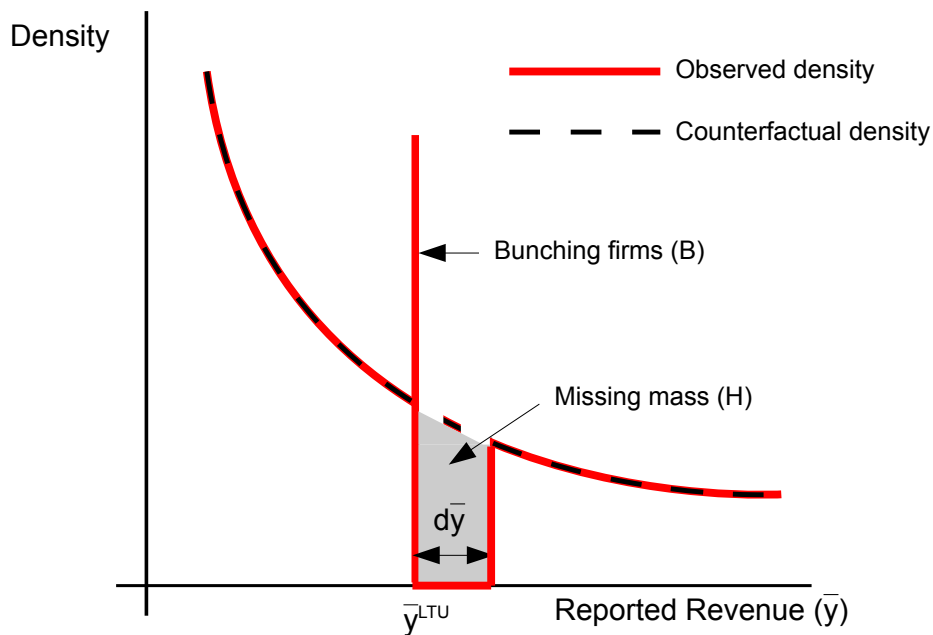
Note: this figure depicts the theoretical revenue distribution before and after the introduction of the Large Taxpayers' Unit (LTU). In the benchmark scenario, all firms face the same probability of detection and the distribution of revenue is smoothly decreasing as depicted by the dashed (black) line. 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 $d\bar{y}$ above the threshold.

Figure 1.2: Operating Revenue Distribution
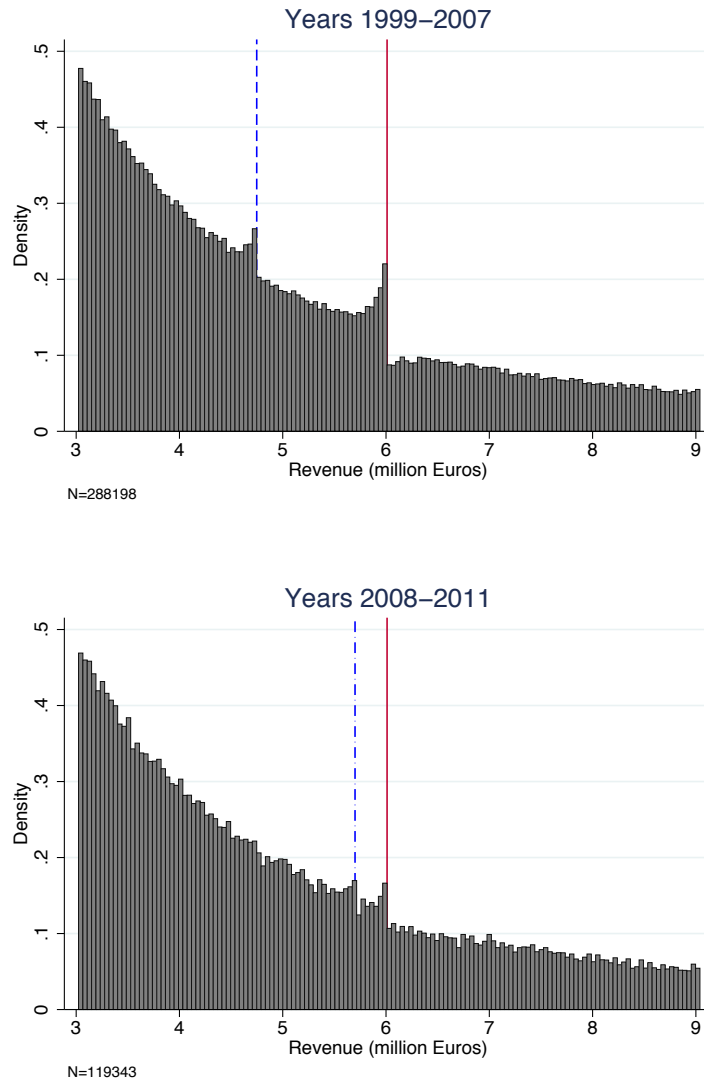
Note: the histograms pool data for the periods 1999-2007 (top) and 2008-2011 (bottom). The dashed (blue) line indicates the External Audit threshold, set at €4.75 million during 1999-2007 and €5.7 million in 2008-2011. The solid (red) line indicates the LTU threshold, set at €6 million for both periods (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 1.3: Revenue Distribution Year by Year, 1999-2007
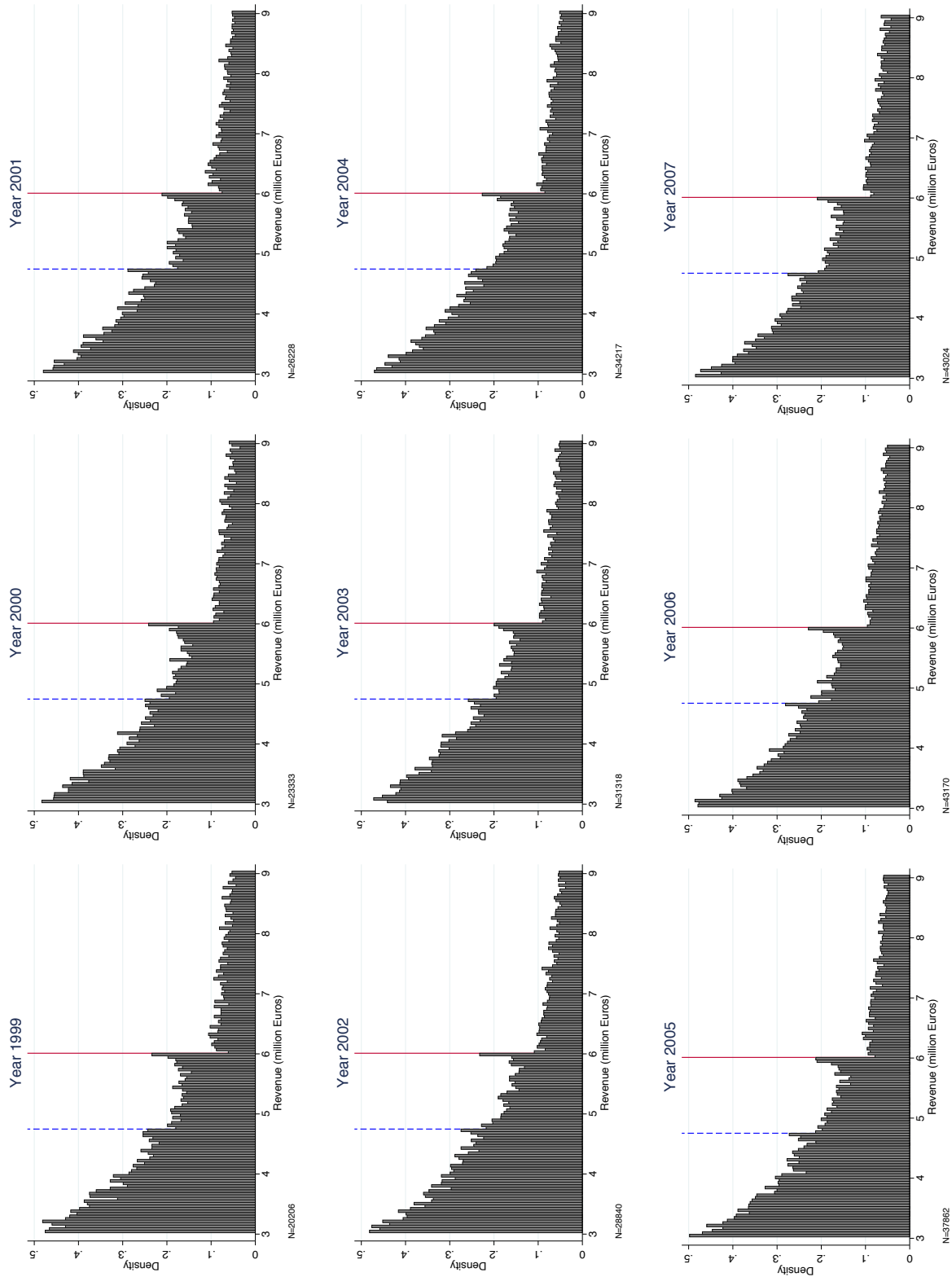
Note: this figure shows annual histograms of reported revenue for each year in the period 1999-2007. The dashed (blue) line indicates the External Audit threshold, set at €4.75 million during 1999-2007. The solid (red) line indicates the LTU threshold, set at €6 million for both periods (and fixed in 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 1.4: Revenue Distribution, Year by Year, 2008-2011
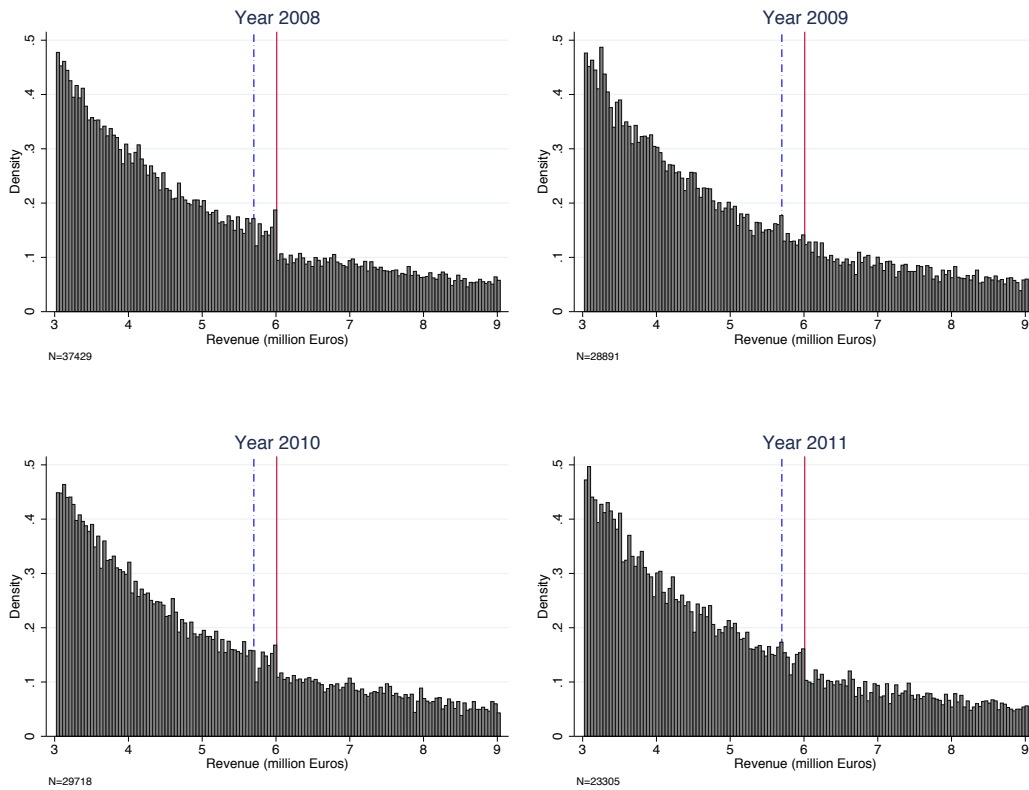
Note: this figure shows annual histograms of reported revenue for each year in the period 2008-2011. The distribution is very similar for all years, with some noise due to the fact that these subsamples are relatively small. The dashed (blue) line indicates the External Audit threshold, set at €5.7 million 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 1.5: Counterfactual Revenue Distribution

Years 1999-2007

b_NF = 0.086 (0.005)

b_F = 0.465 (0.052)

Actual Density   Counterfactual Density

N=291217

Years 2008-2011

b_NF = 0.026 (0.005)

b_F = 0.206 (0.052)

Actual Density   Counterfactual Density

N=121277
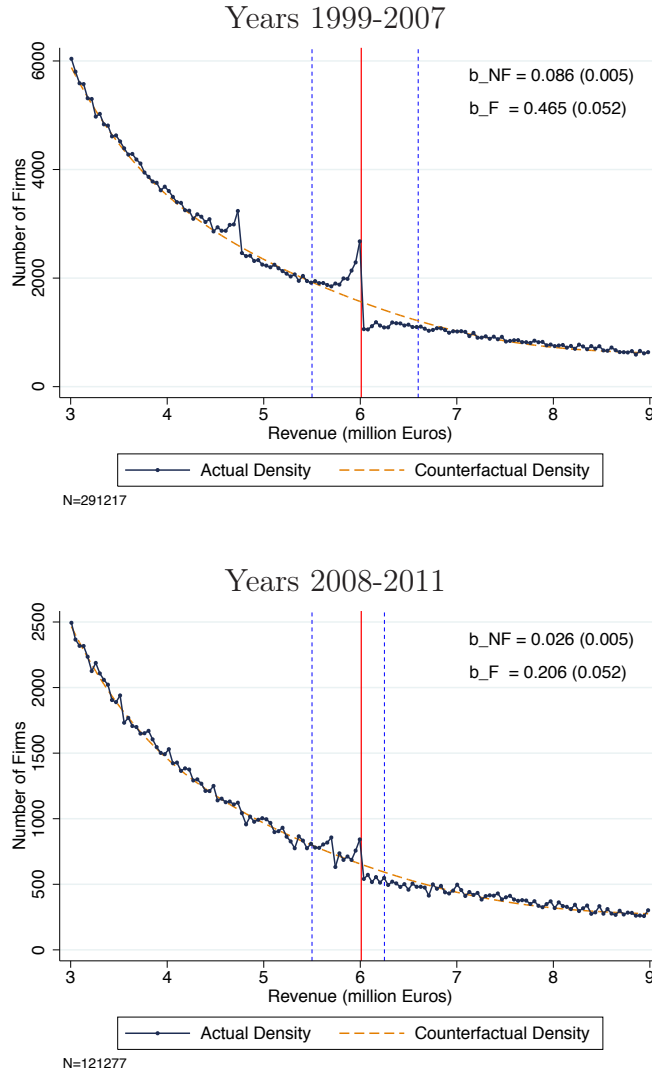
Note: these graphs shows the reported distribution of revenue (dots connected by solid blue line) and the estimated counterfactual (orange dashed curve) for the boom (1999-2007) and recession (2008-2011) periods. The data for the true distribution are exactly the same as that used to construct the histograms in Figure 1.2. 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 1.6: Growing vs. Shrinking Firms

Growing firms
(Revenue_t > Revenue_t−1)

Note: pooled 1999−2011 data
N=242305

Shrinking firms
(Revenue_t < Revenue_t−1)

Note: pooled 1999−2011 data
N=124978
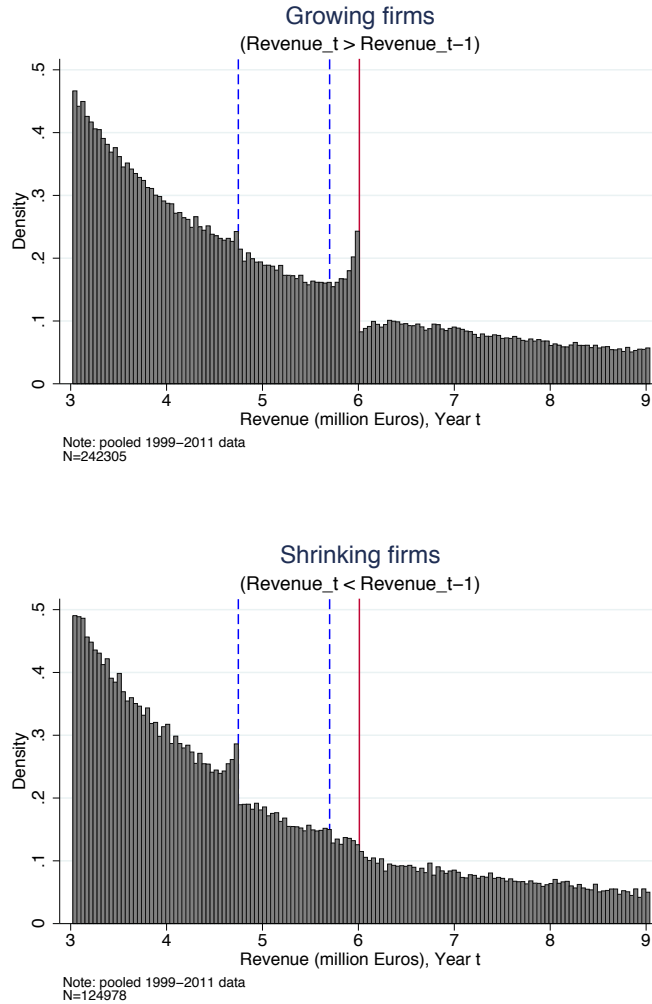
Note: these graphs show annual revenue distributions for two subsamples of firms: those that are growing and those that are shrinking. A firm is defined as growing if its reported revenue in year t is higher than in year $t-1$, i.e. $\overline{y}_t > \overline{y}_{t-1}$. A firm is defined as shrinking 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 1.7: Revenue Distribution by Number of Employees
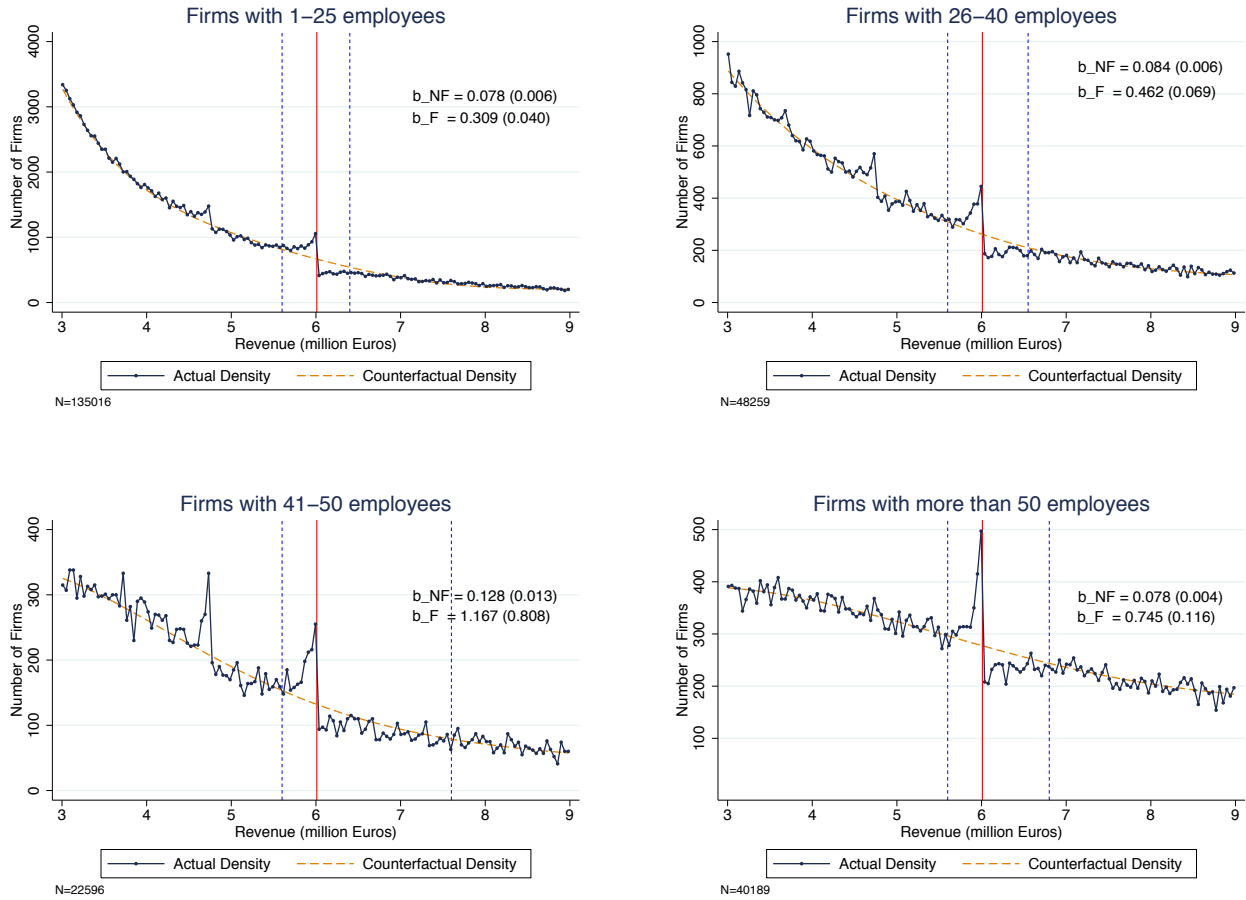
Firms with 1–25 employees

b_NF = 0.078 (0.006)
b_F = 0.309 (0.040)

N=135016

Firms with 26–40 employees

b_NF = 0.084 (0.006)
b_F = 0.462 (0.069)

N=48259

Firms with 41–50 employees

b_NF = 0.128 (0.013)
b_F = 1.167 (0.808)

N=22596

Firms with more than 50 employees

b_NF = 0.078 (0.004)
b_F = 0.745 (0.116)

N=40189

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 1.5. Only data for the period 1999-2007 is used.

Figure 1.8: Revenue Distribution by Fixed Assets
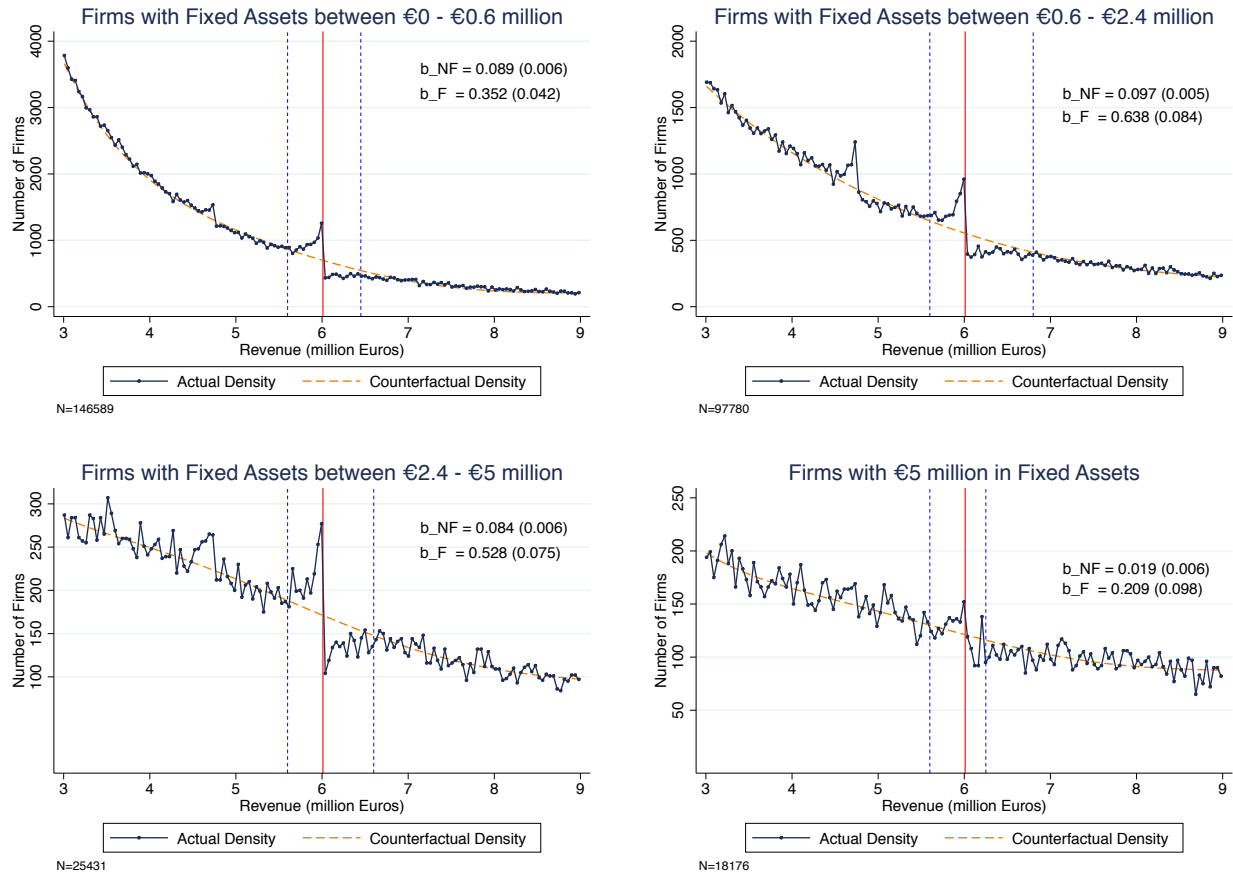
Note: these graphs show the actual and counterfactual revenue distributions for subsamples of firms with a given level of fixed assets (measured in million Euros). The counterfactual distributions are constructed as explained in the note to Figure 1.5. Only data for the period 1999-2007 is used.

Figure 1.9: Revenue Distribution by Organizational Form
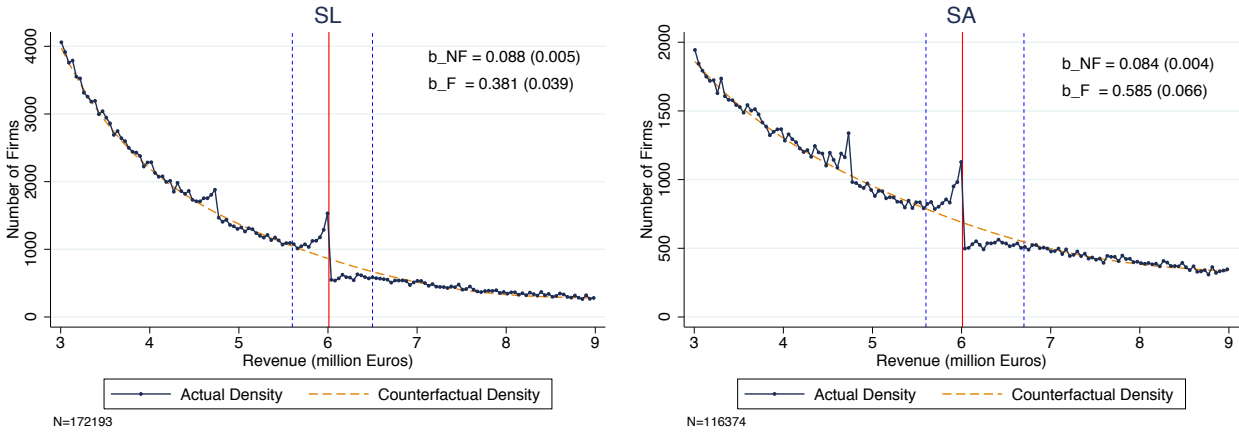
Note: these graphs show the actual and counterfactual revenue distributions for firms with different organizational forms. SL stands for *Sociedad Limited*, equivalent to a Limited Liability Company. SA stands for *Sociedad Anónima*, equivalent to a Corporation. The counterfactual distribution is constructed in each case as explained in the note to Figure 1.5. Only data for the period 1999-2007 is used.

Figure 1.10: 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.

Figure 1.11: Revenue Distribution by Sector of Activity

(a) High Bunching Sectors

Building Contractors

b_NF = 0.132 (0.016)
b_F = 0.395 (0.076)

Number of Firms

Revenue (million Euros)

Actual Density   Counterfactual Density

N=14038

Wholesale Durables

b_NF = 0.112 (0.007)
b_F = 0.549 (0.072)

Number of Firms

Revenue (million Euros)

Actual Density   Counterfactual Density

N=44561
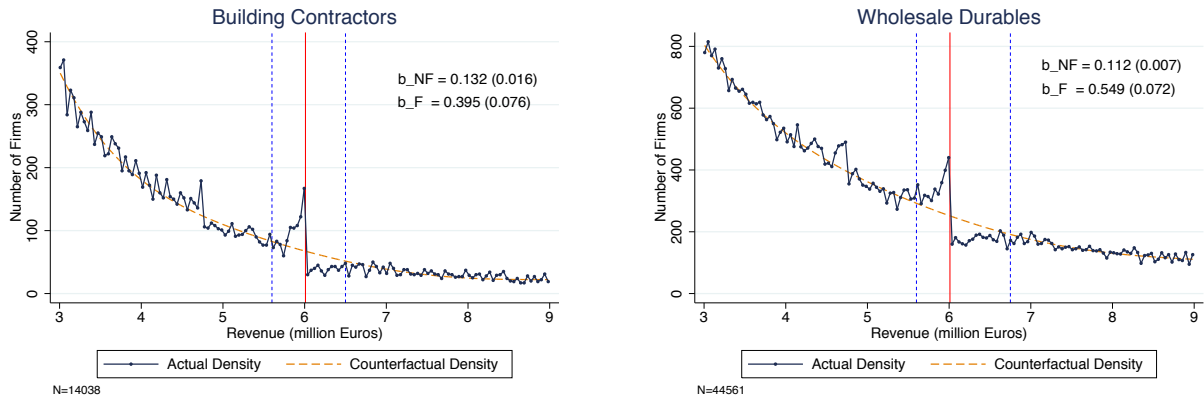
(b) Medium Bunching Sectors

Manufacturing – Metallic Products

b_NF = 0.095 (0.010)
b_F = 0.457 (0.099)

Number of Firms

Revenue (million Euros)

Actual Density   Counterfactual Density

N=19211

Manufacturing – Wood and Paper Products

b_NF = 0.086 (0.009)
b_F = 0.441 (0.085)

Number of Firms

Revenue (million Euros)

Actual Density   Counterfactual Density

N=26408

(c) Low Bunching Sectors

Retail

b_NF = 0.045 (0.011)
b_F = 0.167 (0.053)

Number of Firms

Revenue (million Euros)

Actual Density   Counterfactual Density

N=22411

Restaurants and Hotels

b_NF = 0.023 (0.016)
b_F = 0.336 (2.682)

Number of Firms

Revenue (million Euros)

Actual Density   Counterfactual Density

N=7573

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 1.5. Only data for the period 1999-2007 is used.

Figure 1.12: Bunching Response by Scope of Evasion
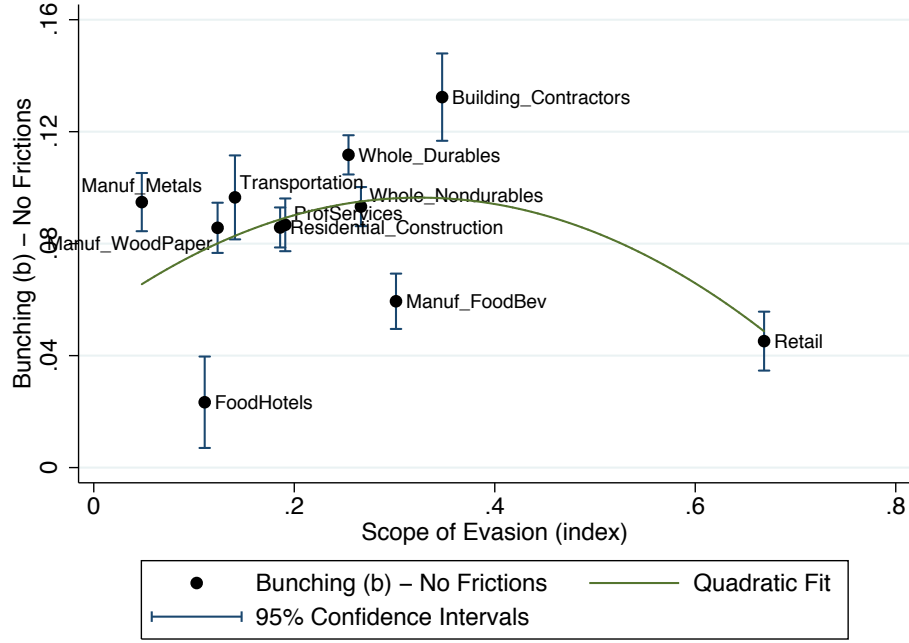
Note: the bunching measure $\widehat{b}_{NF}$ is calculated for each sector as explained in Section 4.1 in the main text. The scope of evasion index is the product of two elements: first, the percentage of a sector's output that is sold to final consumers (obtained from input-output tables published by INE, the National Statistics Institute). Second, the median number of employees that firms with revenue between €5.5-€6.5 million have in a given sector. The intuition for the first element is that selling to final consumers makes underreporting much easier because there VAT self-enforcing mechanism breaks down. For the second element, the idea is that it is easier to underreport if the number of employees is small. Specifically, we compute:

$$Scope = ShareFinalCons * (1 - employees/100)$$

where we divide the number of employees by 100 so that both numbers are smaller than one and $scope \in (0, 1)$.

41

Figure 1.13: Probability of Growth Next Period

Boom Period, 1999-2007

Recession Period, 2008-2011
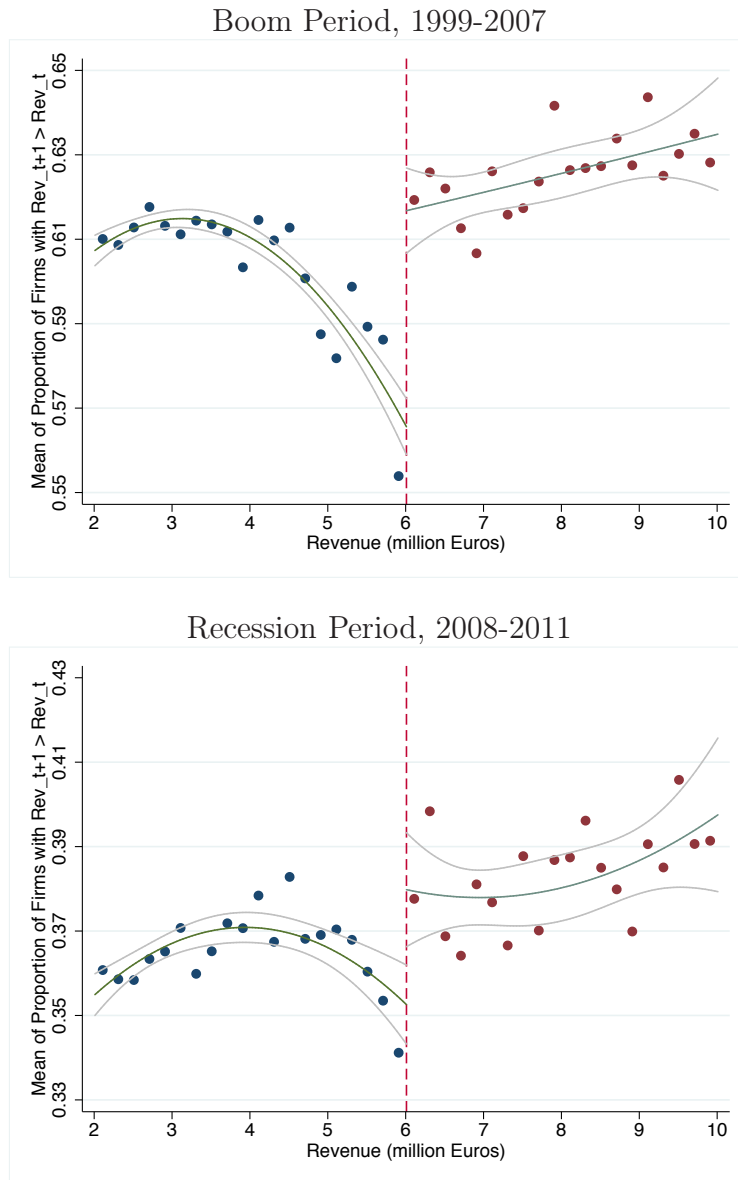
Note: these figures show the probability that firms will report higher revenues the following year, against current reported revenues. The top panel shows results for the period 1999-2007 and the bottom panel for the period 2008-2011. Data are divided in bins of €200,000 each. The dots depict bin averages, while the solid lines are quadratic polynomial fits estimated separately on either side of the threshold. The dashed lines indicated 95% confidence intervals around the quadratic fits.

Figure 1.14: Median Growth Next Period

Boom Period, 1999-2007

Recession Period, 2008-2011
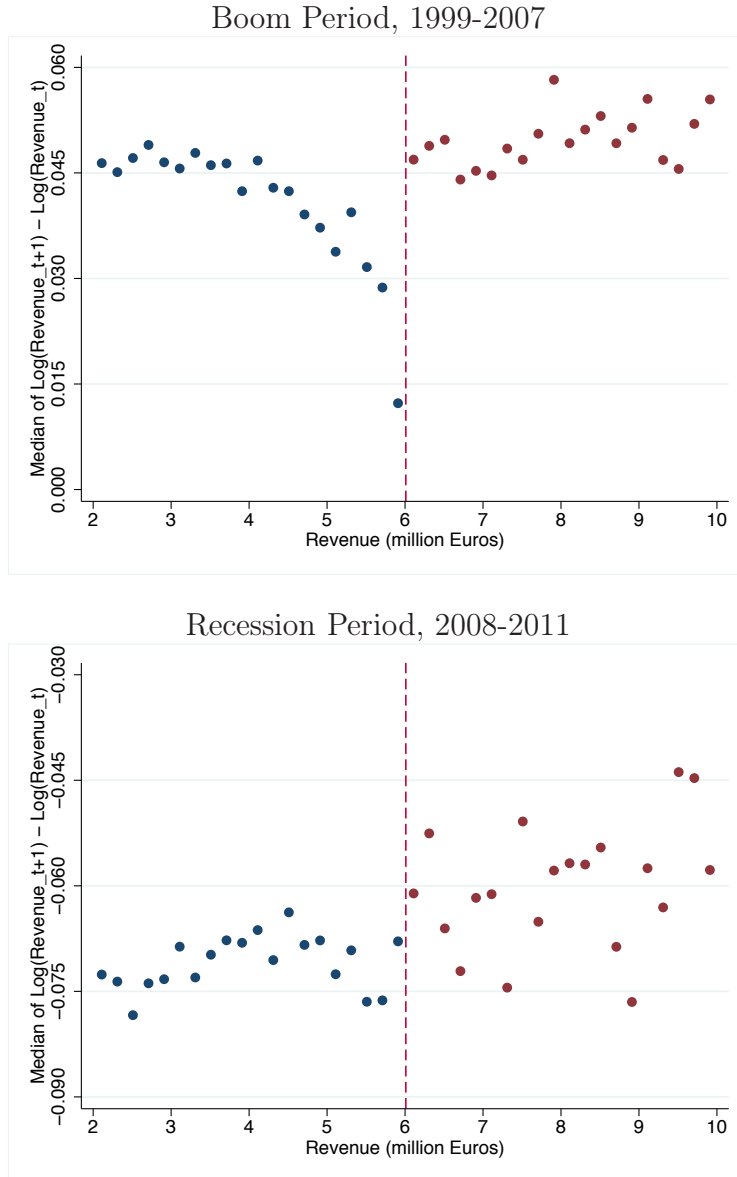
Note: these figures show the median growth in reported revenue experienced by firms in the following year, against current-year reported revenues. The top panel shows results for the period 1999-2007 and the bottom panel for the period 2008-2011. Data are divided in bins of €200,000 each. The dots depict median values for each bin. We do not use averages because they are noisy due to extreme growth values (both positive and negative).

Figure 1.15: Bunching Persistence
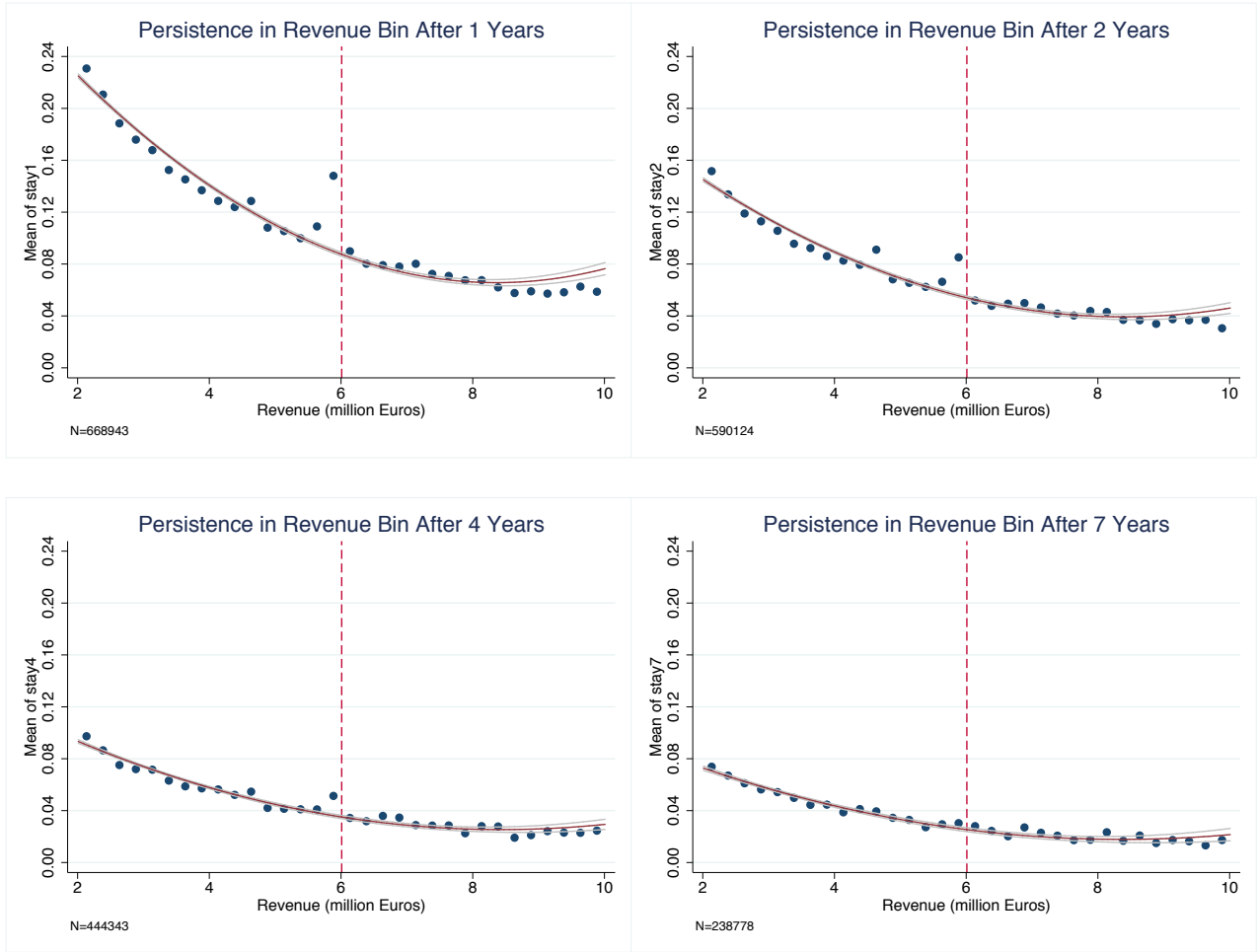
Note: these graphs show the degree of bunching persistence. Data are divided in €250,000 bins such that no bin includes firm both to the left and to the right of the threshold, marked by the dashed (red) vertical line. The blue dots indicate the proportion of firms who reported revenues within that bin 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 $y \in (5.75, 6]$.

Chapter 2

The Impact of Tax Enforcement Policies: Evasion vs. Real Responses and Welfare Implications

with David Lopez-Rodriguez

2.1 Introduction

In Chapter 1, we documented that a set of Spanish firms strategically adjust their reported revenue to remain below the Large Taxpayers' Unit (LTU) threshold. However, an important question remains unanswered: do these firms reduce their actual production (real response) or do they simply underreport their revenues (evasion response)? The distinction between real and evasion responses in both revenue and input expenses is crucial to understand the welfare consequences of the design of the LTU and the incentives it generates.

To make progress in answering this question, we need to extend the stylized model of firm behavior presented in Chapter 1 to capture all potential channels of tax evasion. First, we allow misreporting of input costs, as well as misreporting of revenues. In a model with only one production input, firms can evade the corporate income tax (CIT) by overreporting input costs because it reduces declared profits. We also introduce the value added tax (VAT) in this model, which gives an additional incentive to overreport input costs. In a second extension, we consider a model with two production inputs: labor and materials. This distinction is interesting because the two types of inputs lead to different fiscal incentives: materials are deductible under both the CIT and the 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, depending on the marginal tax rates on the CIT and the PRT. There can be other important incentives to misreport labor costs associated to labor regulations. The uncertainty associated to the production process can create incentives to underreport wage payments to circumvent labor market rigidities.

We use these model extensions to derive testable predictions about the behavior of reported input costs around the LTU threshold. In particular, we analyze how real and evasion

responses would lead to discontinuities at the threshold. If bunching is due to a real response, the model with one input predicts an upward jump of reported input costs at the threshold. This follows from the fact that bunching firms have higher exogenous productivity and the production function is concave, such that bunching firms need fewer inputs to produce the same output as other firms. The same intuition applies to both materials and labor in the model with two production inputs.

If the bunching response is due to tax evasion, the predictions are different. In the model with one input, we would expect a downward jump of reported input costs at the LTU threshold. This is due to bunching firms producing more output (and hence using more inputs) than they report. Moreover, if bunching firms also overreport their inputs, the jump will be even more pronounced. This prediction follows through for material input costs in the model with two inputs. However, the prediction for reported labor input costs is ambiguous: since there are incentives both to under- and overreport them, we cannot say whether they will feature a discontinuity at the threshold.

We bring these predictions to the data by showing plots of the ratio of reported inputs (materials, labor) over revenue in the vertical axis against reported revenue in the horizontal axis. Using ratios rather than levels helps to identify discontinuities at the LTU threshold. We find a 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 are most plausibly due 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. Under the hypothesis of a real response, bunching firms would need to use a smaller number of employees to produce the same output as other firms below the threshold, but there is no significant discontinuity in this variable. To sum up, the graphical evidence seems to rule out the hypothesis that bunching is due entirely to a real response. However, the evidence is not enough to prove that it is all evasion, nor can we disentangle the importance of revenue vs. input cost misreporting.

The graphical analysis described above has several limitations. First of all, it does not allow us to draw causal inference using a standard regression discontinuity approach. This is because firms can manipulate reported revenue –as shown in Chapter 1–, which determines the level of enforcement faced by each firm.[1] Second, the graphs discussed above do not include controls for other firm characteristics or for dynamic aspects of firm behavior. In order to overcome the limitations of the graphical analysis, we exploit the panel structure of the dataset to control for the endogenous sorting around the threshold. We do this by estimating a fixed-effects regression model. The firm fixed effects control for all time invariant characteristics (observable and unobservable) that may drive the sorting process. Identification in this model thus comes from firms that change their enforcement regime

---

[1]Reported revenue is the "assignment variable", as the regression discontinuity literature calls it (Lee and Lemieux, 2010).

during the period under analysis, i.e. they cross from below to above the threshold or vice versa. The dependent variables in the fixed-effects regressions are reported material and labor input costs, number of employees, and average wages.

Once firm fixed effects are included in the regressions, we do not find a significant difference between the reported material and labor input costs on either side of the LTU threshold, contradicting the graphical evidence. This could be due to several factors that bias the fixed-effects coefficients towards zero. The most relevant is that the variation from firms that remain always on the same side of the threshold is absorbed by the fixed effects. Another concern is that firms that go from above the threshold to below may continue to face high enforcement because they are not removed from the LTU census. The behavior of these firms would not change upon crossing the threshold from above to below, biasing the coefficient towards zero.

In order to ease at least the second concern, we define a different indicator of low enforcement that takes value one only if a firm has never reported revenue above the LTU threshold before. In the regressions with this alternative variable, we obtain results that are more similar to the graphical analysis. Firms in low enforcement report material inputs 0.72 percentage points (p.p.) higher than those in high enforcement (compared to an average of 66%) and they report labor costs 0.49 p.p. lower (compared to an average of 13%). Moreover, low enforcement firms report 1.42 fewer employees, but not significantly different average wages. The latter results reverse our earlier findings from the graphical evidence, where most of the lower labor input costs seemed to be due to lower reported wages. Overall, the conclusions from the regression analysis are similar to the ones reached with the graphical evidence: the evasion story seems to be the most plausible explanation of the observed bunching in reported revenue at the LTU threshold (rather than a real production response), but it is hard to pin down whether most of the evasion is due to misreporting of revenue or input costs.

After analyzing the behavioral responses to the discontinuity in tax enforcement, we study the welfare implications of this policy. To do this, we expand on the simple model from Chapter 1 and make two simplifying assumptions. First, we assume that each firm is owned by one individual whose total income is given by the profits of the firm. That way, firm profits can enter a standard utilitarian welfare function. Second, we consider an increase in tax enforcement intensity equivalent to an increase in the expected tax rate. This allows us to draw insights from the extensive literature on the deadweight loss of taxation. Feldstein (1999) made an influential contribution to this literature by establishing that the elasticity of taxable income is a sufficient statistic for the deadweight loss of taxation. Besides the standard labor supply response, his framework allowed for the existence of evasion and avoidance responses, assuming that taxpayers equalize the marginal resource cost of evasion to the tax rate (marginal benefit). Chetty (2009a) pointed out that Feldstein's result does not hold when the marginal *social* cost of evasion is different from the tax rate. This situation arises when taxpayers face a "transfer cost" for evading, such as a monetary penalty if detected. The penalty represents a private cost to the taxpayer, but has no social cost because it is a transfer to the government. Chetty shows that, when taxpayers face transfer costs, the correct sufficient statistic is the elasticity of total income, not just (reported) taxable income. Given that we find little evidence of a real production response, the most natural interpretation of our results is that the efficiency costs of this tax enforcement policy

are likely to be low.

A related question that is interesting from a policy perspective is whether the observed response is relevant in terms of lost tax revenues. Tackling this question is challenging given the limitations of our empirical findings, so we only attempt to provide a rough estimate of the upper bound of corporate income taxes evaded by firms in the low enforcement regime. Making the strong assumption that the total response in reported revenue is actually due to underreporting, we extrapolate the results to all non-LTU firms. This yields estimates of lost tax revenues in the range of 0.17% of GDP (for the "no frictions" estimate) up to 0.95% of GDP (for the "frictions" estimate).

The findings in this paper contribute to the empirical literature on business tax evasion by providing a well identified measure of the effects of tax enforcement on firm behavior. Our finding that stronger enforcement does not create large inefficiencies complements the findings of Paula and Scheinkman (2010) and Pomeranz (2013). These papers emphasize the role of information for effective tax enforcement, particularly in the presence of a VAT that uses the invoice-credit system. The self-enforcing mechanisms of VAT lead to the transmission of evasion (or compliance) behavior up the production chain from retailers to intermediate goods suppliers. Our results indicate that, since the efficiency costs of tax enforcement seem relatively low, an increase in tax enforcement on firms who sell to final consumers could generate positive compliance spillovers upstream in the VAT chain for a low cost.

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[2] (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.[3] This is particularly relevant in the Spanish case, where average firm size is small given the country's level of development.[4] 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.

The rest of the paper is organized as follows. Section 2 provides some institutional background and describes the data. Section 3 presents extensions to the model from Chapter 1. Section 4 presents the graphical analysis and the fixed-effects regressions. Section 5 discusses the welfare implications of the results. Section 6 presents a rough calculation of tax revenue losses. Section 7 concludes.

---

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

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

[4]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.

2.2 Institutional Background and Data

2.2.1 Overview of the Spanish Tax System

The Spanish tax system rests on four main taxes: the payroll tax (PRT),[5] 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 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 (described in Chapter 1). 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.[6] 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.[7] 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.[8]

2.2.2 Data

We use accounting data on Spanish firms from Amadeus for the period 1999-2011. The general characteristics of the dataset are described in Chapter 1. The key variables used in the empirical analysis in this chapter are input costs reported by firms. Two main categories of input costs are reported in Amadeus: materials and labor. The definition of materials

---

[5]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.

[6]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.

[7]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.

[8]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.

includes the monetary cost of raw materials and services purchased by the firm in the production process. Ideally, we would want to analyze services and raw materials separately, but there is no more disaggregation in this dataset. Labor is the total wage bill of the firm. Amadeus does not contain any individual information on employees, so we define an average wage proxy for each firm that is simply the ratio of the total wage bill over the number of employees. In turn, the number of employees is defined as the average number of employees at the end of each month throughout the year. The number of employees is missing, for reasons that we do not know, for about 20% of the firms for which we have information on reported revenue and materials, which reduces the sample size in some of the empirical estimations.

2.3 Theoretical Framework: Extensions

The model in Chapter 1 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.

2.3.1 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:

$$\Pi = p\psi f(m) - cm \tag{2.1}$$

$$\overline{\Pi} = p\left[\psi f(m) - u\right] - c\left[m + v\right] \tag{2.2}$$

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:

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

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:

$$\mathbb{E}\Pi = \Pi - t\overline{\Pi} - \delta t(1+\theta)\left[\Pi - \overline{\Pi}\right] \tag{2.3}$$

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

$$\Pi - \overline{\Pi} = pu + cv \geq 0 \tag{2.4}$$

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

$$(1-t)\,p\psi f_m(m^*) = (1-t)\,c + t(1+\theta)\frac{\partial\delta}{\partial m}\left[pu + cv\right] \tag{2.5}$$

$$tc = t(1+\theta)\left[\frac{\partial\delta}{\partial v}\left[pu + cv*\right] + \delta c\right] \tag{2.6}$$

$$tp = t(1+\theta)\left[\frac{\partial\delta}{\partial u}\left[pu* + cv\right] + \delta p\right] \tag{2.7}$$

Equation (2.5) 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 (2.6), firms equalize the marginal tax savings of overreporting input costs to the marginal expected payment if detected. According to condition (2.7), firms equalize that marginal tax savings of underreporting sales to the marginal expected payment if detected.

Solving (2.6) and (2.7) for v^* and u^* and introducing them into (2.5), we obtain the following expressions:

$$\frac{\partial\delta/\partial v}{\partial\delta/\partial m} = \frac{tc\left[1 - \delta(1+\theta)\right]}{(1-t)\left[p\psi f_m(m^*) - c\right]} \tag{2.8}$$

$$\frac{\partial\delta/\partial u}{\partial\delta/\partial m} = \frac{tp\left[1 - \delta(1+\theta)\right]}{(1-t)\left[p\psi f_m(m^*) - c\right]} \tag{2.9}$$

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 over-reporting and sales underreporting, respectively. Condition (2.8) 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 (2.9) 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:

$$
\begin{aligned}
\Pi &= \left(1 - t^{cit}\right)\left[\left(1 + t^{vat}\right)\left(Y - M\right) - t^{vat}\left(Y - M\right)\right] \\
&= \left(1 - t^{cit}\right)\left[Y - M\right]
\end{aligned}
\tag{2.10}
$$

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:

$$
\begin{aligned}
\Pi^{ND} &= \left(1 - t^{cit}\right)\left\{\left(1 + t^{vat}\right)\left(\overline{Y} - \overline{M}\right) - t^{vat}\left(\overline{Y} - \overline{M}\right)\right\} + \left(U + V\right) \\
&= \left(1 - t^{cit}\right)\left[Y - M\right] + \left(t^{cit} + t^{vat}\right)\left(U + V\right)
\end{aligned}
\tag{2.11}
$$

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:

$$\Pi^D = \left(1 - t^{cit}\right)[Y - M] - (1 + \theta)\left(t^{cit} + t^{vat}\right)(U + V) \tag{2.12}$$

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

$$\mathbb{E}\Pi = \left(1 - t^{cit}\right)[Y - M] + [1 - \delta(1 + \theta)]\left(t^{cit} + t^{vat}\right)(U + V) \tag{2.13}$$

With equation (2.13) 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: $E = U + V$.

2.3.2 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:

$$\begin{aligned}
\Pi &= \left(1 - t^{cit}\right)\left\{\left(1 + t^{vat}\right)[Y - M] - t^{vat}[Y - M] - \left(1 + t^{prt}\right)N\right\} \\
&= \left(1 - t^{cit}\right)\left\{[Y - M] - \left(1 + t^{prt}\right)N\right\}
\end{aligned} \tag{2.14}$$

Equation (2.14) 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:

$$\begin{aligned}
\Pi^{ND} &= \left(1 - t^{cit}\right)\left\{\left(1 + t^{vat}\right)\left[\overline{Y} - \overline{M}\right] - t^{vat}\left[\overline{Y} - \overline{M}\right] - \left(1 + t^{prt}\right)\overline{N}\right\} + (U + V - Z) \\
&= \left(1 - t^{cit}\right)\{Y - M - N\} + \left(t^{cit} + t^{vat}\right)(U + V) + \left[t^{prt} - t^{cit}\left(1 + t^{prt}\right)\right]Z
\end{aligned} \tag{2.15}$$

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

$$\Pi^D = \left(1 - t^{cit}\right)\{Y - M - N\} - \\ - (1 + \theta)\left\{\left(t^{cit} + t^{vat}\right)(U + V) + \left[t^{prt} - t^{cit}\left(1 + t^{prt}\right)\right]Z\right\} \quad (2.16)$$

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

$$\mathbb{E}\Pi = \left(1 - t^{cit}\right)[Y - M - N] + \\ + [1 - \delta(1 + \theta)]\left\{\left(t^{cit} + t^{vat}\right)(U + V) + \left[t^{prt} - t^{cit}\left(1 + t^{prt}\right)\right]Z\right\} \quad (2.17)$$

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 $t^{prt} < t^{cit}\left(1 + t^{prt}\right)$. 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.[9] 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[10] and other countries.[11]

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

---

[9] 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.

[10] 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.

[11] 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.

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 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.

2.3.3 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 Chapter 1 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.

2.4 Empirical Analysis

The predictions presented in the previous section 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. In the second subsection, we present a fixed-effects estimation strategy that attempts to control for sorting behavior to isolate the effect of enforcement on tax reporting.

Recall from Chapter 1 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.

2.4.1 Graphical Evidence

The left panels of Figure 2.1 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.[12] 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.

---

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

The same four plots are shown in Figure 2.2 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.

Figure 2.3 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 2.4 and 2.5 provide a more disaggregated picture of labor input costs. Figure 2.4 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 2.5. 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.

2.4.2   Fixed-Effects Regressions

An alternative method to analyze the behavior of reported inputs under the two enforcement regimes is to use regression analysis. A naïve regression discontinuity approach would com-

pare average input costs reported by firms on either side of the LTU threshold and interpret the differences causally. However, that estimation strategy violates the main assumption underlying regression discontinuity designs because firms can manipulate the assignment variable – in this case, reported revenue.

To deal with the endogenous sorting problem, we estimate a model with firm fixed effects. Controlling for fixed effects teases out any composition effects due to time invariant characteristics, both observable and unobservable, leaving only the variation due to the enforcement regime. However, there are some potential threats to this empirical strategy. Identification of causal effects is based on the behavior of firms that switch their enforcement regime at some point during the period under analysis. Thus, by capturing only within-firm variation, this method may attenuate the estimated effect of enforcement because it cancels out the effects on firms that remain always below (or above) the threshold. We showed in Chapter 1 that there is some persistence in bunching behavior, but this does not fully invalidate the fixed-effects strategy because most firms eventually cross the LTU threshold. We also introduce year dummies to capture year-specific shocks. The model we estimate can be written as follows:

$$z_{it} = \alpha + \phi_i + \lambda_t + \rho \cdot f\left(y_{i\tau} - y^{LTU}\right) + \beta \cdot \mathbb{1}\left[y_{i\tau} \leq y^{LTU}\right] + \varepsilon_{it}, \qquad (2.18)$$

where z_{it} is the expenditure category (material, labor, or all input costs as a percentage of revenue, employees or wages) for firm i in year $t = \{\tau, \tau + 1\}$, ϕ_i denotes firm fixed-effects, λ_t denotes year fixed-effects, $f\left(y_{it} - y^{LTU}\right)$ is a polynomial in the distance to the revenue threshold, and the term $\mathbb{1}\left[y_{it} \leq y^{LTU}\right]$ is an indicator for whether the reported revenue of firm i is in below the LTU threshold in year t.

The way we define the indicator function makes regressions of equation (2.18) equivalent to the graphical evidence shown in the previous section. Under this definition, a firm that crosses from below to above the threshold is treated the same way as a firm that crosses from above to below. According to the law, both directions should have the opposite implication: in the first case the firm is added to the LTU census, in the second it is removed. However, as discussed in Chapter 1, there is an asymmetry and in practice firms remain in high enforcement for some years after crossing back below the threshold. In order to investigate if the "true" enforcement level experienced by firms affects their behavior, we also run regressions where the indicator takes value one only for firms that have never been above the LTU threshold. In this case, the indicator function is $\mathbb{1}\left[y_{it} \leq y^{LTU} \forall t \leq \tau\right]$. We label the first definition "Below threshold" and the second "Low enforcement".

As we did in the graphical analysis, we present results for the outcome variables in year t and also year $t+1$, to account for the possibility that firm behavior only responds to changes in the level of enforcement with a lag. We use the entire period of data available, 1999-2011, because the year dummies control for aggregate time effects. Finally, in all regressions we cluster standard errors at the firm level to address potential serial correlation.

The regression results using the different input categories as dependent variable are reported in Tables 2.2-2.6. Columns (1)-(4) report the results from the contemporaneous regressions (outcome variable and enforcement indicator are defined in the same year τ), while in columns (5)-(8) the outcome variable is measured in year $\tau + 1$. Odd-numbered columns do not include firm fixed effects and even-numbered columns include them. This

affects mainly the R-squared of regressions, which is much higher when fixed effects are included.

Table 2.2 reports the results when the dependent variable is overall reported input costs. The coefficient on the "below threshold" dummy is -0.35 percentage points (p.p.) when fixed effects are not included (column 1), but the sign is reversed to 0.18 p.p. when fixed effects are included. Both coefficients are statistically significant at the 1% level. When considering "low enforcement", the coefficient is positive and also close to 0.2 p.p., and significant when fixed effects are included (column 4). The results when the dependent variable is measured in the following year (columns 5-8) are close to zero and insignificant. Since the mean of the dependent variable is around 92%, as indicated by the constant in column (1), the economic significance of all these estimates is small.

Table 2.3 reports the regression results for material input costs. When we consider "below threshold" as the treatment indicator, we find a significant coefficient of 1.67 p.p. in column (1), broadly consistent with the graphical evidence. The coefficient is also above one when the outcome variable is measured in year $t+1$, but fixed effects reduce the estimates close to zero. When considering only firms actually under low enforcement as treated, however, the coefficient is 0.72 p.p. (from an average of 66%) and significant. The coefficient is half the size for materials in year $t+1$, but also significant. These results indicate that reported material inputs are higher among low-enforcement firms after controlling for time invariant characteristics and year effects.

The results for labor input costs are reported in Table 2.4. As was the case for materials, the size of the coefficients shrinks to zero when firm fixed effects are included, but they are still statistically significant when the treatment variable is "low enforcement". The coefficient is -0.49 p.p. for labor inputs for labor in year t and -0.80 p.p. for labor in year $t+1$ (average labor input costs are approximately 16% or revenue). The conclusion is that firms facing low enforcement report substantially lower labor inputs even after including firm fixed effects.

Tables 2.5 and 2.6 report the results for number of employees and average wages. In the case of number of employees, all the coefficients in the fixed effects regressions are negative and significant for both treatment variables. Specifically, firms facing low enforcement report 1.42 fewer employees than those facing high enforcement (column 4). In the case of wages, there is a substantial difference between the coefficients with and without fixed effects. According to the regressions without fixed effects, firms in low enforcement pay wages 1,944 euros lower (column 3) or 2,115 euros lower in the following year (column 7). This is broadly consistent with the graphical evidence. When the fixed effects are included, the effect vanishes and the coefficient is close to zero and insignificant. Therefore, it seems like one of the results from the graphical evidence is reversed here: the lower labor input costs reported by firms facing low enforcement would be due to a smaller average number of employees, rather than lower reported wages.

2.5   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:

$$\mathbb{EW}\left(\delta\right) = \left\{\left(1 - t\right)\Pi + tu\left[1 - \delta\left(1 + \theta\right)\right]\right\} + t\left[\Pi - u\left[1 - \delta\left(1 + \theta\right)\right]\right] \qquad (2.19)$$

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:

$$\frac{d}{d\delta}\mathbb{EW} = t\left[-u\left(1 + \theta\right) + \frac{d\Pi}{d\delta} - \frac{du}{d\delta}\left[1 - \delta\left(1 + \theta\right)\right] + u\left(1 + \theta\right)\right]$$

$$= t\frac{d\Pi}{d\delta} - t\left[\frac{du}{d\delta}\left[1 - \delta\left(1 + \theta\right)\right]\right] \qquad (2.20)$$

$$= t\frac{d\overline{\Pi}}{d\delta} + t\frac{du}{d\delta}\left[\delta\left(1 + \theta\right)\right] \qquad (2.21)$$

We know from comparative statics that $\frac{du}{d\delta}|_{u=u^*} < 0$, so the second term in (2.21) 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,

$$t\frac{d\Pi}{d\delta} \leq \frac{d}{d\delta}\mathbb{EW} \leq t\frac{d\overline{\Pi}}{d\delta}. \qquad (2.22)$$

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 (2009a) 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 Chapter 1 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 (2.20) and (2.21) 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 (2.20) and (2.21) would be:

$$\frac{d}{d\delta}\mathbb{E}W = t\frac{d\Pi}{d\delta} - t\left[\frac{du}{d\delta}\left[1 - \delta\left(1 + \theta\right)\right]\right] - \frac{dq}{d\delta} \tag{2.23}$$

$$= t\frac{d\overline{\Pi}}{d\delta} + t\frac{du}{d\delta}\left[\delta\left(1 + \theta\right)\right] - \frac{dq}{d\delta} \tag{2.24}$$

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).[13] 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-

---

[13]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).

benefit calculation to determine what is the socially optimal enforcement intensity, we need to consider the tax revenue lost due to low enforcement.

2.6 Estimates of Lost Tax Revenue

In Chapter 1, we showed that the marginal buncher reduces its reported revenue by 1.4% in the case of no optimization frictions, and up to 7.5% once frictions are taken into account. These effects are precisely estimated and statistically significant. The empirical evidence presented in this chapter suggest that most of the bunching response is due to revenue underreporting rather than a real production response. Putting aside for a moment the potential misreporting of inputs, we perform a rough estimation of the upper bound of evasion in the corporate income tax associated to the bunching behavior.

To perform these calculations, we make a number of assumptions. First, we extrapolate our local estimates for firms near €6 million in revenue to all firms with smaller reported revenue. This implies assuming that firms below the LTU threshold conceal the same percentage of their revenue as the marginal buncher. Concealing revenue is in theory easier for small firms because they have simpler operations and fewer employees than large firms. Thus, we think this assumption makes the extrapolation of local results acceptable. Second, we assume that the marginal corporate income tax rate is 30% for all non-LTU firms (even though the actual rate was 35% for some of these firms in the early years of the 1999-2007 period).

The thought experiment we propose is the following: what would be the tax liability of non-LTU firms if they reported their true revenue? We define true revenue as actual reported revenue plus the percentage underreported, according to our estimates from Chapter 1. We use data from Amadeus for the year 2005 to perform the calculations (the results are similar for other years). In 2005, the corporate income tax raised €38 billion (4.2% of GDP), of which €8.3 billion came from non-LTU firms, according to official statistics from AEAT (Several years). There are 553,956 firms in Amadeus with revenue below €6 million, of which 64.8% report positive profits. In the official statistics, there are 1.1 million firms in that revenue range, of which 42% report positive profits. Applying the statutory 30% tax rate on the firms with positive profits in Amadeus and summing over firms yields a total tax liability of €7.98 billion. The number is quite close to the official statistics, despite the fact that Amadeus has only half the number of firms. This indicates that the firms missing from Amadeus are mostly small firms with little or no declared profits.

We extrapolate the "true" revenue of each firm using our two estimates (no frictions and frictions), and make the assumption that inputs were reported truthfully. The results are reported in Table 2.7. The proportion of firms reporting positive profits rises to 75.1% (in the case of no frictions) and up to 83.4% (in the frictions case). The increase in overall tax revenue from the corporate income tax is €1.5 billion (0.17% of GDP) using the no frictions estimate, and €8.5 billion (0.95% of GDP) using the estimate that accounts for frictions.

These numbers are reasonably large considering that they almost double (in the frictions case) the total tax revenue raised from non-LTU firms in practice. To put these results in perspective, keep in mind that we interpret this number as the upper bound of tax evasion due to revenue misreporting. Considering the extent of fiscal consolidation currently facing

the Spanish public sector (close to 10% of GDP), recovering up to one percentage point of GDP through stricter tax enforcement (itself a costly activity) does not appear to be a solution by itself.

2.7   Conclusions and Future Work

This paper has tried to disentangle whether firms in Spain respond to a discontinuity in tax enforcement intensity created by the Large Taxpayers' Unit by reducing production (real response) or by misreporting revenue and input costs (evasion response). We have tested the predictions of an extended model with two input costs (materials and labor) and three taxes (corporate income tax, payroll tax and VAT). Both the graphical evidence and fixed-effects regressions suggest that a fully real response is not compatible with the patterns in the data, so it can be ruled out. However, it is not possible to describe the type of evasion response because the empirical strategies cannot disentangle to what extent firms misreport their revenue and their input costs.

The broad welfare implication of these results is that this type of tax enforcement policy does not generate large efficiency losses because it only affects reporting, not real activity. This in turn, leads to the question of whether the evasion response could lead to a substantial loss of tax revenues for the government. A rough calculation considering only the corporate income tax put the upper bound of tax revenue losses at 0.95% of GDP, which is a relatively low figure given the strong assumptions made to compute it.

We intend to continue analyzing this policy discontinuity with an improved dataset compiled by the Bank of Spain. This new dataset uses the same original source as Amadeus (the original financial statements submitted by firms to the Commercial Registry). The advantage is that the new dataset does not drop any information from the reports, which Amadeus does in order to obtain homogeneous data for all European countries. Thanks to this, the new dataset contains information about firms' exporting activity, which is useful because exporters are always included in the LTU and therefore unaffected by the eligibility threshold. Using exporting firms as a comparison group is a promising avenue of research. The Bank of Spain's dataset also has a finer level of disaggregation for input costs, which will also be useful to deepen the empirical analysis.

Tables

Table 2.1: 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 2.2: All Input Costs. Fixed-Effects Regressions, 1999-2011

	All Input Costs (% of Revenue), year τ				All Input Costs (% of Revenue), year $\tau+1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Below Threshold ($y_{i\tau} \leq y^{LTU}$)	-0.354*** (0.146)	0.181* (0.094)			0.288* (0.120)	0.020 (0.061)		
Low Enforcement ($y_{it} \geq y^{LTU} \forall t \leq \tau$)			0.117 (0.121)	0.210*** (0.080)			-0.217 (0.134)	-0.130 (0.103)
Constant	92.082*** (0.148)	92.322*** (0.101)	92.280*** (0.098)	89.931*** (0.090)	92.800*** (0.004)	93.400*** (0.133)	93.139*** (0.111)	89.800*** (0.114)
Firm Fixed Effects	no	yes	no	yes	no	yes	no	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
Quadratic in Revenue	yes	yes	yes	yes	yes	yes	yes	yes
Observations	263,953	263,953	263,953	263,953	263,953	263,953	263,953	263,953
Clusters	64,994	64,994	64,994	64,994	64,994	64,994	64,994	64,994
R-squared	0.00	0.80	0.00	0.81	0.00	0.74	0.01	0.74

Note: Standard errors are clustered at the firm level. Significance levels: *** = 1%, ** = 5%, and * = 10%. The sample is restricted to firms that reported revenues between €3-€9 million at least one year in the period 1999-2011.
In columns (1), (2), (5), and (6), we estimate the following model:

$$z_{it} = \alpha + \phi_i + \lambda_t + \rho \cdot f\left(y_{it} - y^{LTU}\right) + \beta \cdot \mathbb{1}\left[y_{i\tau} \leq y^{LTU}\right] + \varepsilon_{it},$$

where z_{it} is total reported inputs as a percentage of total revenue for firm i in year $t = \{\tau, \tau + 1\}$, ϕ_i denotes firm fixed-effects, λ_t denotes year fixed-effects, $f\left(y_{it} - y^{LTU}\right)$ is a quadratic polynomial in the distance of reported revenue to the threshold, and $\mathbb{1}\left[y_{i\tau} \leq y^{LTU}\right]$ is an indicator for whether firm i's revenue is below the LTU threshold in year τ.
In columns (3), (4), (7) and (8), we estimate the model:

$$z_{it} = \alpha + \phi_i + \lambda_t + \rho \cdot f\left(y_{it} - y^{LTU}\right) + \beta \cdot \mathbb{1}\left[y_{it} \geq y^{LTU} \forall t \leq \tau\right] + \varepsilon_{it},$$

where the indicator $\mathbb{1}\left[y_{it} \leq y^{LTU} \forall t \leq \tau\right]$ takes value one for firms that effectively experience low enforcement, i.e., firms whose reported revenue has never been above the LTU threshold.

65

Table 2.3: Material Input Costs. Fixed-Effects Regressions, 1999-2011

	Material Input Costs (% of Revenue), year τ				Material Input Costs (% of Revenue), year $\tau+1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Below Threshold $(y_{i\tau} \leq y^{LTU})$	1.675***	0.059			1.229***	-0.213		
	(0.339)	(0.114)			(0.350)	(0.152)		
Low Enforcement $(y_{it} \leq y^{LTU} \, \forall t \wedge t \leq \tau)$			1.679***	0.720***			0.693**	0.372***
			(0.312)	(0.104)			(0.317)	(0.130)
Constant	66.502***	67.658***	66.723***	66.985***	65.658***	64.813***	66.122***	64.401***
	(0.341)	(0.128)	(0.258)	(0.115)	(0.349)	(0.164)	(0.261)	(0.139)
Firm Fixed Effects	no	yes	no	yes	no	yes	no	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
Quadratic in Revenue	yes	yes	yes	yes	yes	yes	yes	yes
Observations	312,481	312,481	312,481	312,481	312,481	312,481	312,481	312,481
Clusters	75,694	75,694	75,694	75,694	75,694	75,694	75,694	75,694
R-squared	0.01	0.94	0.01	0.94	0.01	0.91	0.01	0.91

Note: Standard errors are clustered at the firm level. Significance levels: *** $= 1\%$, ** $= 5\%$, and * $= 10\%$. The sample is restricted to firms that reported revenues between €3-€9 million at least one year in the period 1999-2011. In columns (1), (2), (5), and (6), we estimate the following model:

$$z_{it} = \alpha + \phi_i + \lambda_t + \rho \cdot f\left(y_{it} - y^{LTU}\right) + \beta \cdot \mathbb{1}\left[y_{i\tau} \leq y^{LTU}\right] + \varepsilon_{it},$$

where z_{it} is material input costs as a percentage of total revenue for firm i in year $t = \{\tau, \tau + 1\}$, ϕ_i denotes firm fixed-effects, λ_t denotes year fixed-effects, $f\left(y_{it} - y^{LTU}\right)$ is a quadratic polynomial in the distance of reported revenue to the threshold, and $\mathbb{1}\left[y_{i\tau} \leq y^{LTU}\right]$ is an indicator for whether firm i's revenue is below the LTU threshold in year τ. In columns (3), (4), (7) and (8), we estimate the model:

$$z_{it} = \alpha + \phi_i + \lambda_t + \rho \cdot f\left(y_{it} - y^{LTU}\right) + \beta \cdot \mathbb{1}\left[y_{it} \leq y^{LTU} \, \forall t \leq \tau\right] + \varepsilon_{it},$$

where the indicator $\mathbb{1}\left[y_{it} \leq y^{LTU} \forall t \leq \tau\right]$ takes value one for firms that effectively experience low enforcement, i.e., firms whose reported revenue has never been above the LTU threshold.

Table 2.4: Labor Input Costs. Fixed-Effects Regressions, 1999-2011

	Labor Input Costs (% of Revenue), year τ				Labor Input Costs (% of Revenue), year $\tau + 1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Below Threshold $(y_{i\tau} \leq y^{LTU})$	-0.682*** (0.208)	0.040 (0.060)			-0.557** (0.218)	0.026 (0.098)		
Low Enforcement $(y_{it} \leq y^{LTU} \forall t \leq \tau)$			-1.298*** (0.194)	-0.491*** (0.058)			-1.129*** (0.202)	-0.804*** (0.084)
Constant	18.842*** (0.209)	12.351*** (0.074)	18.999*** (0.148)	13.054*** (0.070)	20.065*** (0.219)	14.567*** (0.112)	20.306*** (0.156)	15.707*** (0.095)
Firm Fixed Effects	no	yes	no	yes	no	yes	no	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
Quadratic in Revenue	yes	yes	yes	yes	yes	yes	yes	yes
Observations	316,924	316,924	316,924	316,924	316,924	316,924	316,924	316,924
Clusters	75,730	75,730	75,730	75,730	75,730	75,730	75,730	75,730
R-squared	0.01	0.96	0.01	0.96	0.02	0.90	0.02	0.90

Note: Standard errors are clustered at the firm level. Significance levels: *** = 1%, ** = 5%, and * = 10%. The sample is restricted to firms that reported revenues between €3–€9 million at least one year in the period 1999-2011. In columns (1), (2), (5), and (6), we estimate the following model:

$$z_{it} = \alpha + \phi_i + \lambda_t + \rho \cdot f \left(y_{it} - y^{LTU} \right) + \beta \cdot \mathbb{1} \left[y_{i\tau} \leq y^{LTU} \right] + \varepsilon_{it},$$

where z_{it} is labor input costs as a percentage of total revenue for firm i in year $t = \{\tau, \tau + 1\}$, ϕ_i denotes firm fixed-effects, λ_t denotes year fixed-effects, $f \left(y_{it} - y^{LTU} \right)$ is a quadratic polynomial in the distance of reported revenue to the threshold, and $\mathbb{1} \left[y_{i\tau} \leq y^{LTU} \right]$ is an indicator for whether firm i's revenue is below the LTU threshold in year τ. In columns (3), (4), (7) and (8), we estimate the model:

$$z_{it} = \alpha + \phi_i + \lambda_t + \rho \cdot f \left(y_{it} - y^{LTU} \right) + \beta \cdot \mathbb{1} \left[y_{it} \leq y^{LTU} \forall t \leq \tau \right] + \varepsilon_{it},$$

where the indicator $\mathbb{1} \left[y_{it} \leq y^{LTU} \forall t \leq \tau \right]$ takes value one for firms that effectively experience low enforcement, i.e., firms whose reported revenue has never been above the LTU threshold.

Table 2.5: Number of Employees. Fixed-Effects Regressions, 1999-2011

	Number of Employees, year τ				Number of Employees, year $\tau+1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Below Threshold ($y_{i\tau} \le y^{LTU}$)	-1.427***	-0.394**			-1.533***	-0.501**		
	(0.495)	(0.198)			(0.506)	(0.226)		
Low Enforcement ($y_{it} \le y^{LTU} \forall t \le \tau$)			-1.755***	-1.423***			-0.701	-0.742***
			(0.445)	(0.185)			(0.447)	(0.208)
Constant	37.739***	34.861***	37.653***	36.039***	39.168***	35.669***	38.441***	36.002***
	(0.517)	(0.242)	(0.391)	(0.216)	(0.532)	(0.263)	(0.399)	(0.231)
Firm Fixed Effects	no	yes	no	yes	no	yes	no	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
Quadratic in Revenue	yes	yes	yes	yes	yes	yes	yes	yes
Observations	271,704	271,704	271,704	271,704	271,704	271,704	271,704	271,704
Clusters	71,227	71,227	71,227	71,227	71,227	71,227	71,227	71,227
R-squared	0.05	0.92	0.05	0.93	0.05	0.91	0.05	0.91

Note: Standard errors are clustered at the firm level. Significance levels: *** = 1%, ** = 5%, and * = 10%. In columns (1), (2), (5), and (6), we estimate the following model:

$$z_{it} = \alpha + \phi_i + \lambda_t + \rho \cdot f\left(y_{it} - y^{LTU}\right) + \beta \cdot \mathbb{1}\left[y_{i\tau} \le y^{LTU}\right] + \varepsilon_{it},$$

where z_{it} is the number of employees in firm i in year $t = \{\tau, \tau+1\}$, ϕ_i denotes firm fixed-effects, λ_t denotes year fixed-effects, $f\left(y_{it} - y^{LTU}\right)$ is a quadratic polynomial in the distance of reported revenue to the threshold, and $\mathbb{1}\left[y_{i\tau} \le y^{LTU}\right]$ is an indicator for whether firm i's revenue is below the LTU threshold in year τ. In columns (3), (4), (7) and (8), we estimate the model:

$$z_{it} = \alpha + \phi_i + \lambda_t + \rho \cdot f\left(y_{it} - y^{LTU}\right) + \beta \cdot \mathbb{1}\left[y_{it} \le y^{LTU} \forall t \le \tau\right] + \varepsilon_{it},$$

where the indicator $\mathbb{1}\left[y_{it} \le y^{LTU} \forall t \le \tau\right]$ takes value one for firms that effectively experience low enforcement, i.e., firms whose reported revenue has never been above the LTU threshold.

Table 2.6: Average Wages. Fixed-Effects Regressions, 1999-2011

	Average Wages (thousands of euros), year τ				Average Wages (thousands of euros), year $\tau + 1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Below Threshold ($y_{i\tau} \leq y^{LTU}$)	-0.841*** (0.234)	-0.146 (0.165)			-0.829*** (0.244)	-0.173 (0.176)		
Low Enforcement ($y_{it} \leq y^{LTU} \forall t \leq \tau$)			-1.944*** (0.206)	-0.256 (0.165)			-2.115*** (0.211)	-0.287* (0.167)
Constant	26.656*** (0.225)	23.931*** (0.183)	27.435*** (0.156)	24.317*** (0.166)	27.774*** (0.231)	24.749*** (0.193)	28.651*** (0.159)	25.082*** (0.169)
Firm Fixed Effects	no	yes	no	yes	no	yes	no	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
Quadratic in Revenue	yes	yes	yes	yes	yes	yes	yes	yes
Observations	268,885	268,885	268,885	268,885	268,885	268,885	268,885	268,885
Clusters	70,398	70,398	70,398	70,398	70,398	70,398	70,398	70,398
R-squared	0.04	0.78	0.05	0.78	0.04	0.77	0.04	0.77

Note: Standard errors are clustered at the firm level. Significance levels: *** $= 1\%$, ** $= 5\%$, and * $= 10\%$. In columns (1), (2), (5), and (6), we estimate the following model:

$$z_{it} = \alpha + \phi_i + \lambda_t + \rho \cdot f\left(y_{it} - y^{LTU}\right) + \beta \cdot \mathbb{1}\left[y_{i\tau} \leq y^{LTU}\right] + \varepsilon_{it},$$

where z_{it} is average wages (in thousands of euros) in firm i in year $t = \{\tau, \tau + 1\}$, ϕ_i denotes firm fixed-effects, λ_t denotes year fixed-effects, $f\left(y_{it} - y^{LTU}\right)$ is a quadratic polynomial in the distance of reported revenue to the threshold, and $\mathbb{1}\left[y_{i\tau} \leq y^{LTU}\right]$ is an indicator for whether firm i's revenue is below the LTU threshold in year τ. In columns (3), (4), (7) and (8), we estimate the model:

$$z_{it} = \alpha + \phi_i + \lambda_t + \rho \cdot f\left(y_{it} - y^{LTU}\right) + \beta \cdot \mathbb{1}\left[y_{it} \leq y^{LTU} \forall t \leq \tau\right] + \varepsilon_{it},$$

where the indicator $\mathbb{1}\left[y_{it} \leq y^{LTU} \forall t \leq \tau\right]$ takes value one for firms that effectively experience low enforcement, i.e., firms whose reported revenue has never been above the LTU threshold.

Table 2.7: Lost Tax Revenue Calculations

	Amadeus	**No frictions**	**Frictions**
Revenue Measure:	*As reported*	*Reported + 1.4%*	*Reported + 7.5%*
Observations	553,956	553,956	553,956
% with Positive Profits	64.8%	75.1%	83.4%
Average Tax Liability (million €)	0.014	0.017	0.030
Total Tax Liability (million €)	7,988.5	9,508.5	16,578.9
Difference (million €)	–	1,520.0	8,590.4
Difference (% of tax revenue)	–	0.48%	2.65%
Difference (% of GDP)	–	0.17%	0.95%

Note: this table summarizes the calculations of lost tax revenue in the low enforcement regime for the year 2005. The first column shows the actual observations from Amadeus. The next two columns present the results of creating a new reported revenue measure equal to actual reported revenue plus a percentage based on the bunching estimates: 1.4% (no frictions estimate), and 7.5% (frictions estimate).
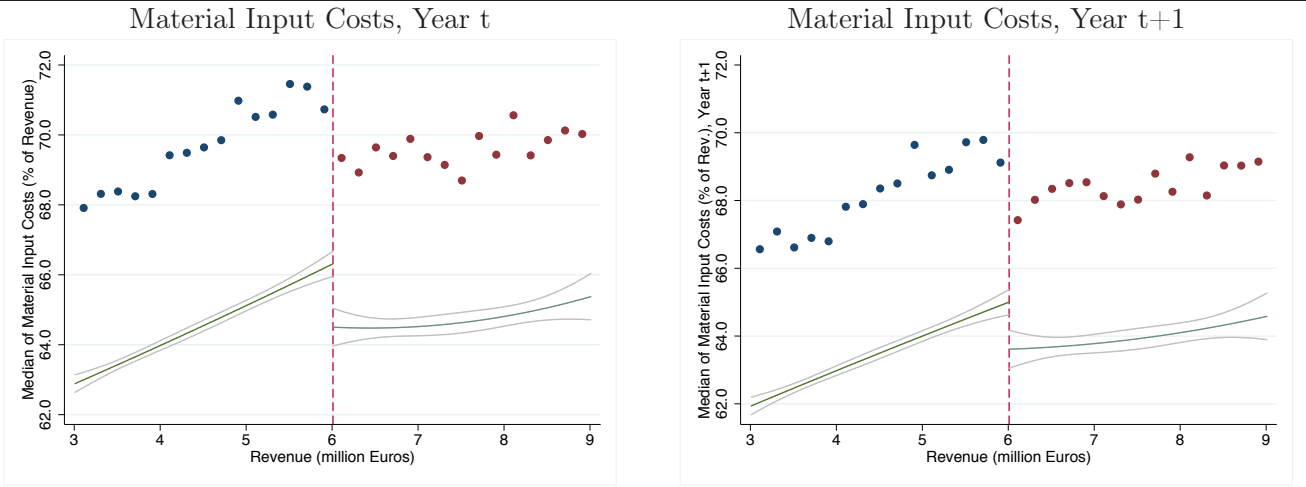
Figures

Figure 2.1: Reported Input Costs

Boom Period, 1999-2007
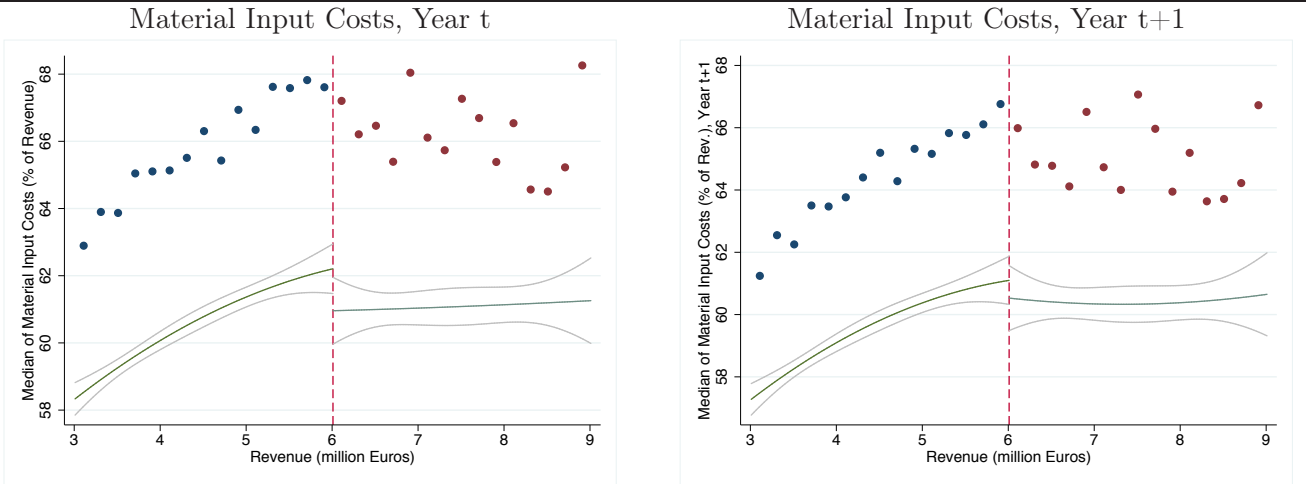

Recession Period, 2008-2011
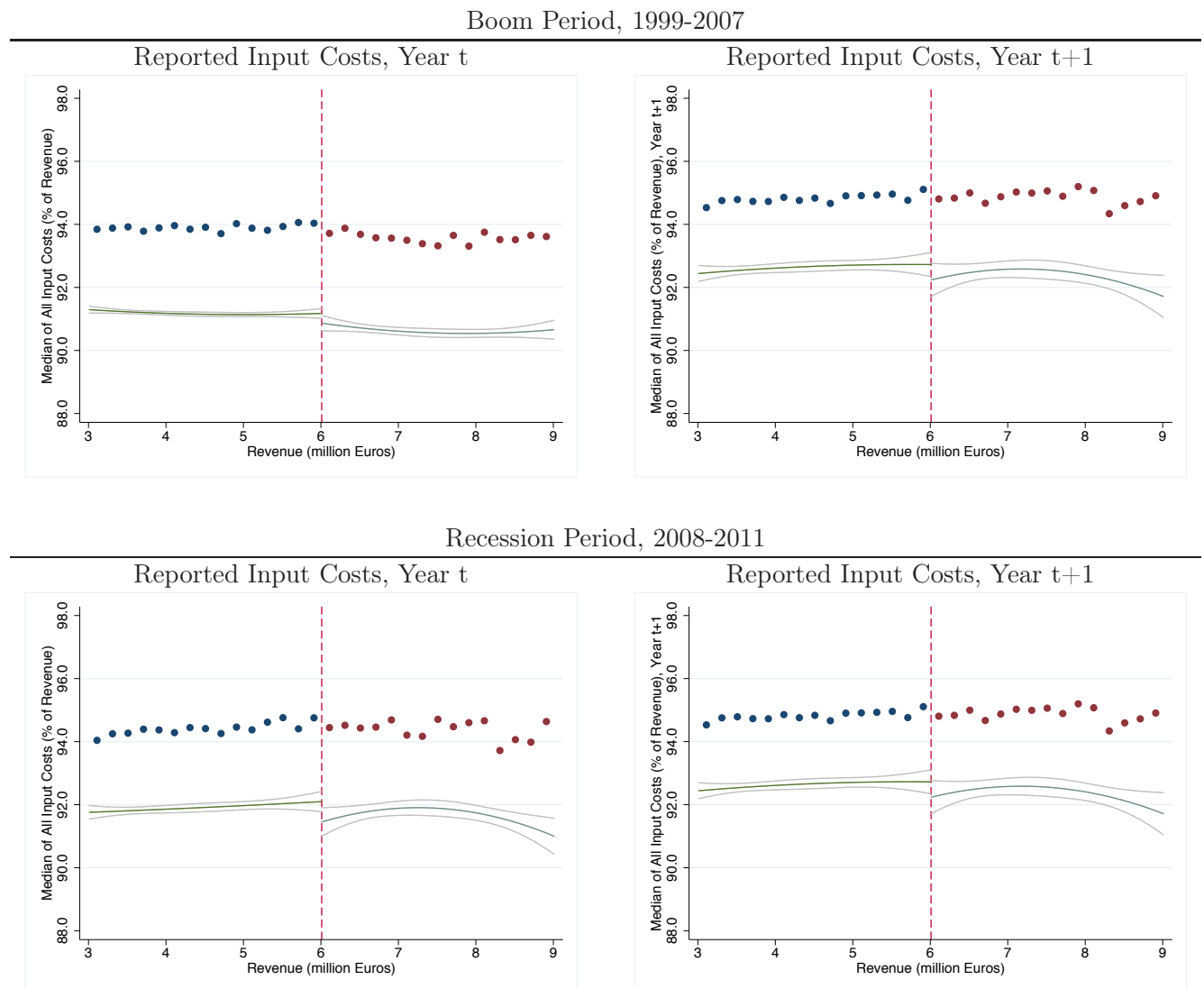
Note: these graphs show the patterns of the ratio of reported input costs over revenue for the period 1999-2007 (top panels) and 2008-2011 (bottom panels). Data are divided in €250,000 bins such that no bin includes firms both to the left and to the right of the LTU threshold, which is marked by the dashed (red) vertical line. The dots depict bin medians. The solid curves represent quadratic polynomial fits estimated separately on each side of the threshold, and the dashed curves are 95% confidence intervals.

Figure 2.2: Material Input Costs
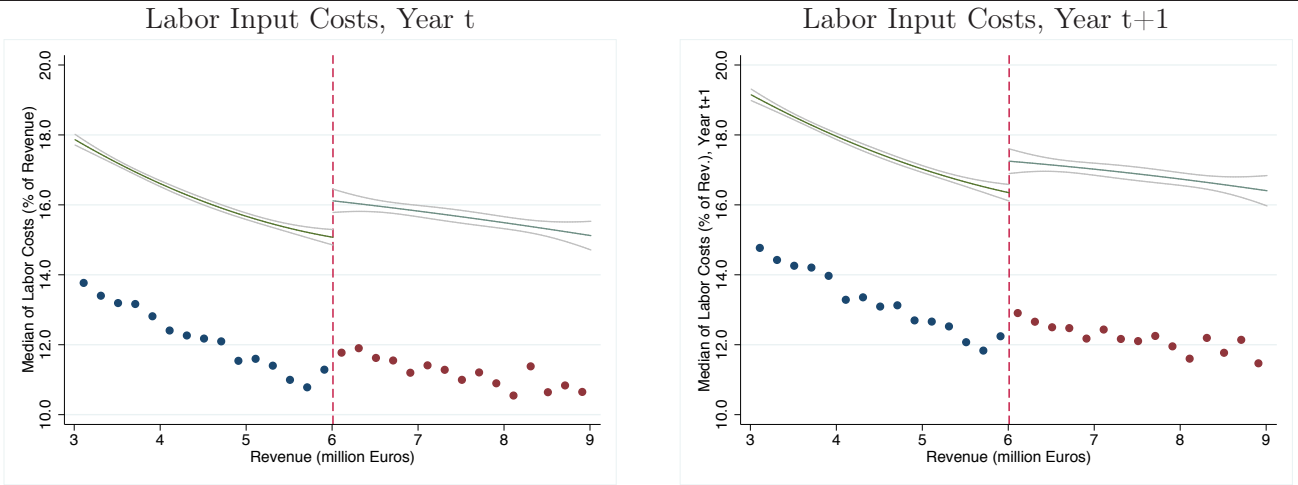
Boom Period, 1999-2007
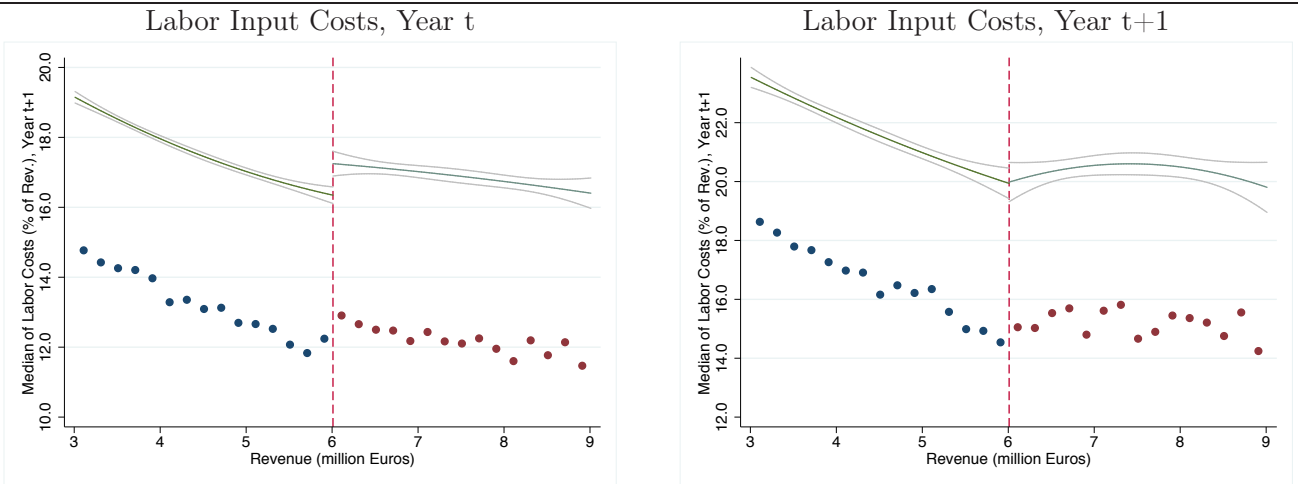

Recession Period, 2008-2011

Note: these graphs show the patterns of the ratio of material input costs over revenue for the period 1999-2007 (top panels) and 2008-2011 (bottom panels). 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 medians. The solid curves represent quadratic polynomial fits estimated separately on each side of the threshold, and the dashed curves are 95% confidence intervals.

Figure 2.3: Labor Input Costs
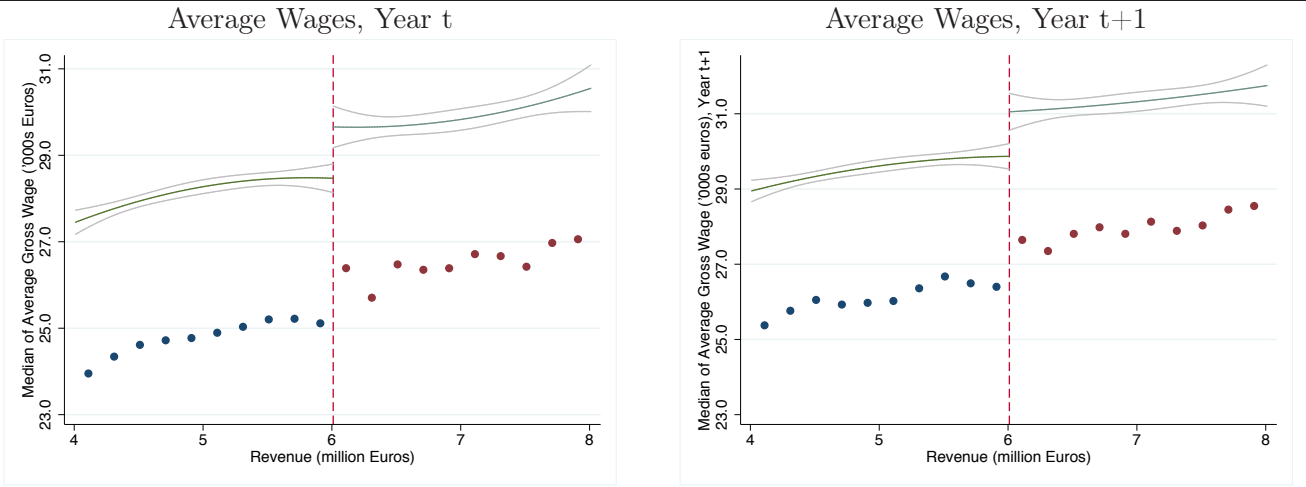
Boom Period, 1999-2007

Labor Input Costs, Year t

Labor Input Costs, Year t+1

Recession Period, 2008-2011

Labor Input Costs, Year t

Labor Input Costs, Year t+1

Note: these graphs show the patterns of the ratio of labor input costs over revenue for the period 1999-2007 (top panels) and 2008-2011 (bottom panels). 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 medians. The solid curves represent quadratic polynomial fits estimated separately on each side of the threshold, and the dashed curves are 95% confidence intervals.

Figure 2.4: Average Wages

Boom Period, 1999-2007

Average Wages, Year t

Average Wages, Year t+1

Recession Period, 2008-2011

Average Wages, Year t

Average Wages, Year t+1
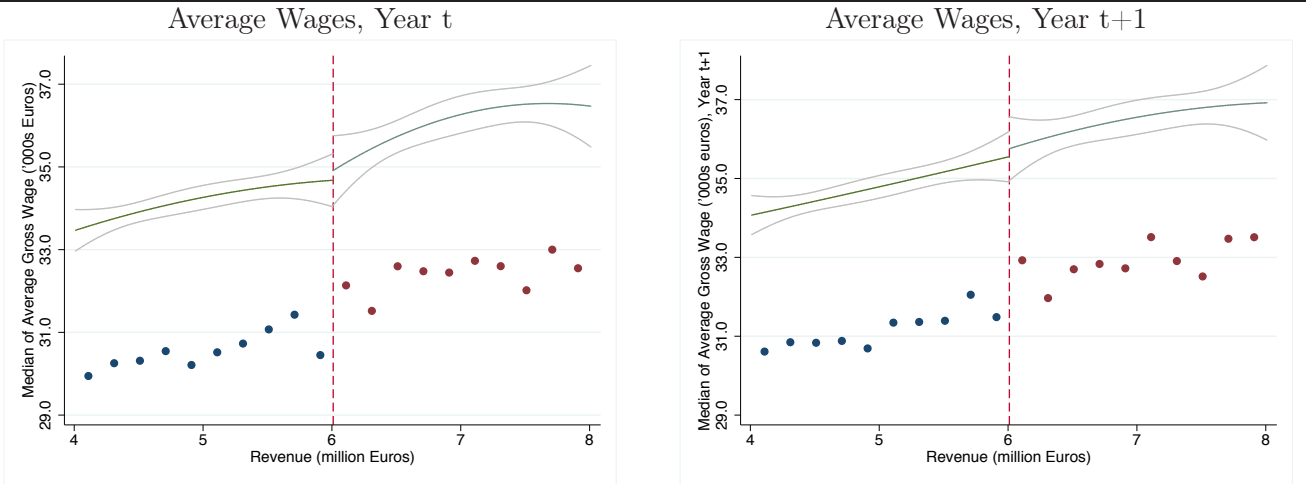
Note: these graphs show the patterns of average wages (in thousands of euros) for the period 1999-2007 (top panels) and 2008-2011 (bottom panels). Data are divided in €250,000 bins such that no bin includes firms both to the left and to the right of the LTU threshold, which is marked by the dashed (red) vertical line. The dots depict bin medians. The solid curves represent quadratic polynomial fits estimated separately on each side of the threshold, and the dashed curves are 95% confidence intervals.

Figure 2.5: Number of Employees
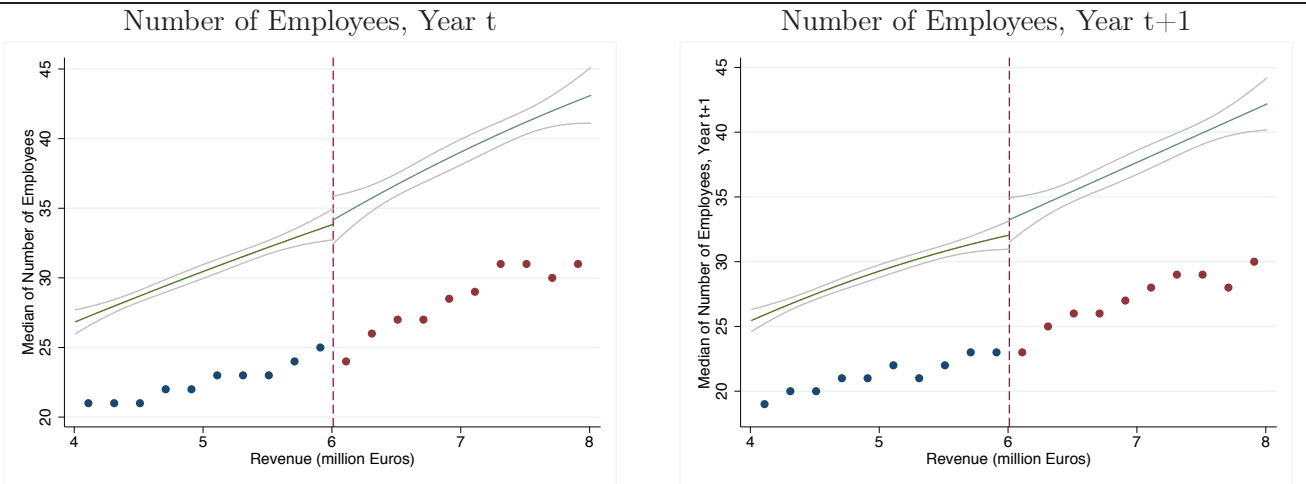
Boom Period, 1999-2007
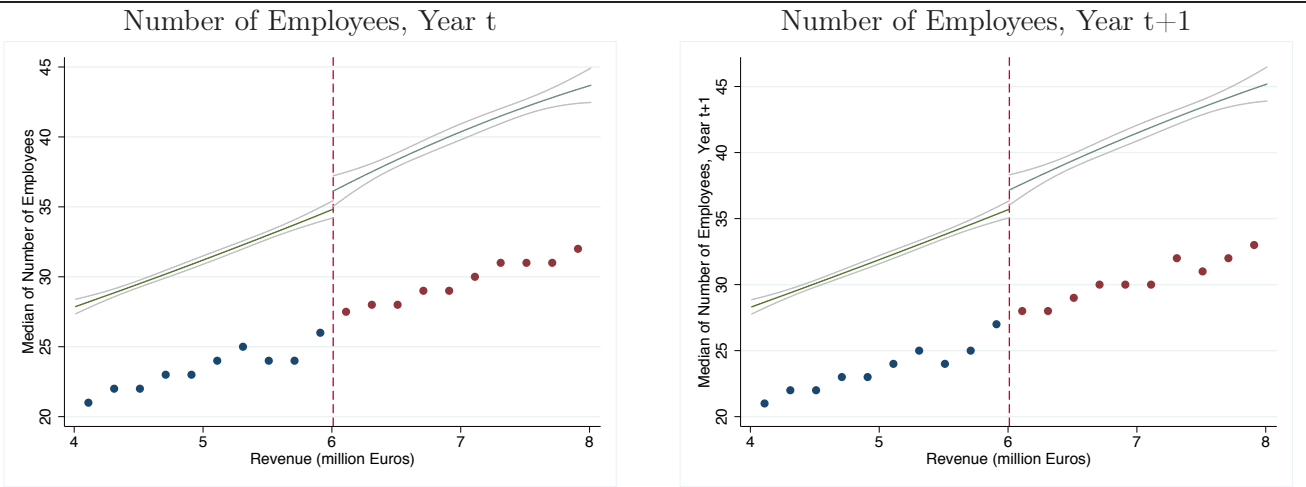
Recession Period, 2008-2011

Note: these graphs show the patterns of the number of employees for the period 1999-2007 (top panels) and 2008-2011 (bottom panels). 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 medians. The solid curves represent quadratic polynomial fits estimated separately on each side of the threshold, and the dashed curves are 95% confidence intervals.

Chapter 3

Firm Size Responses to Labor Regulations: Evidence from France

3.1   Introduction

Size-contingent labor regulations are ubiquitous in many countries. The goal of such regulations is often to give small firms more labor market flexibility and increase their growth potential. However, the existence of regulatory thresholds dampens the incentives to grow, potentially defeating the purpose of these policies (Heckman and Pages, 2003). Understanding the impact of this type of regulations on employment and firm growth is important to inform the design of labor market policies.

In this paper, I study the impact of a regulatory threshold in France, where there are 34 different labor laws that apply only to firms with 50 or more employees. One of these laws requires firms to create a *comité d'entreprise* (works council) that empowers unions within the firm; another imposes bureaucratic limits to worker separations, and a third one establishes of a profit-sharing scheme by which employees receive a share of annual profits.[1] Overall, these regulations increase the average labor cost per employee for French businesses by about 4% (according to Attali, 2008) and give firms a strong incentive to remain below the 50-employee threshold.

In the absence regulatory thresholds, we would expect the firm size distribution, measured by the number of employees, to be smoothly decreasing as in other developed economies (see, for example, Braguinsky, Branstetter and Regateiro, 2011). Using data from Amadeus for the period 2002-2008, I document that a nontrivial group of firms strategically choose to "bunch" below the threshold in order to avoid the additional regulations. The bunching pattern is stable when analyzing each annual sample separately, so I rely on the pooled dataset to obtain more precise estimates. In terms of growth patterns, I show that the proportion of firms increasing their size from one year to the next drops almost by half at 49 employees (35% vs. 20%), while the share of firms keeping a constant number of employees doubles at 49 employees (15 vs. 30%).

The structural parameter of interest in this context is the elasticity of labor demand

---

[1]Ceci-Renaud and Chevalier (2010) and the online appendix of Garicano, LeLarge and van Reenen (2013) provide detailed descriptions of all the relevant regulations.

implied by the response to the regulatory threshold. Departing from the traditional labor economics approach surveyed by Hamermesh (1986, 1993), I apply techniques from a recent literature in public finance that uses discontinuities in tax systems to estimate income tax elasticities (Saez, 2010; Chetty et al., 2011; Kleven and Waseem, 2013).[2] To estimate the elasticity of labor demand, I first set up a stylized model with labor as the only factor of production and a discontinuity in average labor costs at an employment cutoff. In this model, firms receive an exogenous productivity draw (as in Lucas, 1978) that determines their optimal size. The model predicts the existence of three types of firms in equilibrium: (i) low-productivity firms that are too small to be affected by regulations, (ii) bunching firms that cut back their number of employees to be exactly at the cutoff, and (iii) high-productivity firms that reduce their number of employees due to the regulations, but not enough to avoid them.

I derive an expression for the elasticity of labor demand that can be estimated using the number of bunching firms as a sufficient statistic.[3] The key to the estimation strategy is to construct a counterfactual firm size distribution and then estimate the number of bunching firms as the difference between the observed distribution and the counterfactual in an interval below the cutoff. Applying this method, I obtain a point estimate of $e_{NF} = 0.055$ for the elasticity of labor demand pooling data from the period 2002-2008, which is statistically different from zero at the 10% level. Making an adjustment for the possibility that some firms do not respond to the regulations due to optimization frictions, I obtain a point estimate of $e_F = 0.572$, which can be interpreted as an upper bound for the long-term structural elasticity, although it is imprecisely estimated (the standard error is 0.668).

These point estimates are considerably below labor demand elasticities estimated in the literature, which according to Hamermesh (1993) tend to be in the interval $(0.15, 0.75)$. An intuitive explanation for why I obtain such a low point estimate is that bunching firms may be adjusting their production by increasing the use of other inputs instead of labor. I find some preliminary evidence supporting this hypothesis: median fixed assets per employee drop sharply at the notch, indicating that bunching firms have a higher capital-labor ratio than firms just above the threshold. This implies that the model should incorporate other production inputs such as capital, and the estimation strategy should consider the elasticity of substitution between capital and labor. This is left for future work.

The study of French labor regulations and their influence on the firm size distribution is not new. Ceci-Renaud and Chevalier (2010) were the first to document the bunching at the 50-employee regulatory threshold in a descriptive study where they also document bunching at 10 and 20 employees.[4] Garicano, LeLarge and van Reenen (2013) focus on the impact of these regulations on the productivity of French firms. They structurally estimate

---

[2]Saez (2010) and Chetty et al. (2011) study "kinks" in the budget set, define as points where the marginal tax rate jumps discontinuously. Meanwhile, Kleven and Waseem (2013) study "notches", defined as points where *both* the marginal and the average tax rate jump. In the current paper, I study a notch in labor costs, because there is a discontinuous upward jump in labor costs at the 50-employee threshold. For a review of the relationship between the two concepts, see Slemrod (2010).

[3]Chetty (2009b) summarizes the general characteristics of this approach, which he argues is a bridge between reduced-form and structural estimation methods.

[4]Schivardi and Torrini (2008) provide similar evidence for Italy, where key employment protection legislation applies only to firms with more than 15 employees.

a model based on Lucas (1978) and calculate that the welfare loss associated with labor market regulations is 0.8% of GDP assuming flexible wages and up to 5.1% of GDP when assuming wage rigidities. The key contribution of the current paper is to provide an estimate for the labor demand elasticity using an alternative empirical approach to the same setting. Knowledge about this structural parameter can be useful to understand firm behavior in response to other shock and also inform new labor market policies.

This paper is also related to a literature that shows how micro-level distortions can affect aggregate productivity through resource misallocation. Guner, Ventura and Xu (2008) calibrate a growth model with endogenous size distribution of establishments. They find that size-dependent policies can generate substantial reductions in output by increasing the equilibrium number of establishments but reducing output per establishment.[5] Hsieh and Klenow (2009) show that resource misallocation accounts for a significant share of the differences in aggregate productivity between China, India and the US.

The remainder of the paper is organized as follows. Section 2 presents the theoretical framework and derives an estimation strategy for the labor demand elasticity. Section 3 provides background on French labor regulations and describes the data. Section 4 reports the graphical evidence and the elasticity estimations. Section 5 discusses the welfare implications. Section 6 concludes.

3.2 Theoretical Framework and Empirical Strategy

3.2.1 Setup

Consider an economy with a finite number of firms and no entry or exit. Each firm has a managerial productivity θ_i with p.d.f. $g(\theta)$, and a production function where labor (l) is the only factor of production, where $F'(l) > 0$ and $F''(l) < 0$.[6] Labor is infinitely elastically supplied at wage rate w. All non-wage labor costs stemming from labor regulations are combined into a single parameter, φ, that I will refer to as a "labor tax". Notice that φ is expressed as a percentage of net wages received by workers, w. Let $C(l, \varphi, w)$ denote the cost function faced by the firm, which I assume to be linear: $C(l, \varphi, w) = (1 + \varphi) wl$.

In the standard version of the model, labor costs are the same for all firms, $\varphi_i = \varphi^0 \; \forall i$. Firms maximize the following profit function:

$$\max_{\{l_i\}} \pi_i \quad = \quad \theta_i F(l_i) - (1 + \varphi^0) wl_i \tag{3.1}$$

The first order condition for an interior solution is

$$l_i(\theta_i, \varphi^0) = F'^{-1}\left(\frac{(1 + \varphi^0) w}{\theta_i}\right), \tag{3.2}$$

which defines optimal labor demand for firm i. Since we have assumed that $\theta_i \sim g(\theta)$, there exists some $h(\cdot)$ such that $l_i(\theta_i, \varphi) \sim h(l)$, where the shape of $h(\cdot)$ depends on $g(\cdot)$ and

---

[5]In a closely related paper, Restuccia and Rogerson (2008) build a growth model with heterogeneous establishments to show that policy distortions (defined in a general way that includes size-dependent policies) that apply only to some of them make the most efficient ones produce too little and employ too few workers.

[6]The basic setup of this model follows the seminal contribution of Lucas (1978) to the literature of firm size distributions, although that model does not incorporate taxes.

the specific production function. If the p.d.f. of the managerial productivity distribution is smooth, the distribution of firm size will also be smooth.

3.2.2   Introducing a Small Notch

The government introduces a new set of labor regulations that increase the labor tax from φ^0 to $\varphi^1 > \varphi^0$. These regulations only apply to firms above a certain number of employees l^*, that is:

$$\varphi_i = \begin{cases} \varphi^0 & \text{if } l_i \leq l^* \\ \varphi^1 = \varphi^0 + \Delta\varphi & \text{if } l_i > l^* \end{cases} \tag{3.3}$$

where $\Delta\varphi > 0$ is assumed to be small. From the point of view of the firm, the new regulations increase the average cost per employee, including those that already worked for the firm (i.e., all inframarginal employees). Hence, this changes introduces a notch in the labor cost function at l^* employees. A notch is defined as a point in the budget set where both the average and the marginal tax rate jump. It is different from a kink, which is a point where the marginal tax rate jumps but the average tax rate varies smoothly.

Figure 3.1 depicts the incentives created by the notch. The straight lines from the origin represent labor cost functions under the two labor taxes, φ^0 and φ^1. After the reform, firms face the labor cost function represented by the solid sections of these lines, with a discontinuous upward jump at the notch. The graph shows isoprofit curves for two firms, A and B, with managerial productivity levels θ_A and $\theta_B > \theta_A$. Firm A maximizes profits choosing $l_A = l^*$ under the initial labor tax rate and also under the higher rate. In contrast, Firm B chooses $l_B(\varphi^0) > l^*$ under the low labor tax but is indifferent between choosing l^* and $l^* + dl^*$ employees once the notch is introduced. I assume, without loss of generality, that firm B chooses $l_B(\varphi^1) = l^*$. It follows that every firm i with managerial productivity $\theta_i \in (\theta_A, \theta_B)$ also chooses $l_i = l^*$.

The consequences of this analysis for the firm size distribution (measured by the number of employees) can be seen in Figure 3.2, which depicts the theoretical firm size distribution (measured by number of employees). The counterfactual density distribution is depicted by the black dashed curve, which is smoothly decreasing by the assumptions on the distribution of managerial productivity. The observed distribution, depicted in the red solid curve, depicts what happens once the notch is introduced. There are three types of firms in the new equilibrium with the notch: low productivity firms, bunching firms, and high productivity firms. Low productivity firms have $\theta_i \in [\theta_{min}, \theta^*)$ and they choose $l_i < l^*$ under both tax regimes.[7] Bunching firms have productivities in the range $\theta_i \in [\theta^*, \theta^* + d\theta^*)$, and they all choose $l_i = l^*$ in response to the reform, "bunching" at the notch.[8] I define firm B as the marginal buncher, i.e. the firm with the highest managerial productivity level that chooses $l_i = l^*$. Finally, high productivity firms have $\theta_i \in [\theta^* + d\theta^*, \theta_{max})$ and they choose $l_i > l^*$ both with and without the notch. These firms reduce their demand for employment after the notch is introduced, but they do not bunch at the notch because it is too costly for them to reduce their scale that much.

---

[7]For completeness, firms with $\theta_i < \theta_{min}$ never enter the market, because their optimal labor demand is negative so they cannot produce anything.

[8]Notice that $\theta_A = \theta^*$ and $\theta_B = \theta^* + d\theta^*$.

3.2.3   Derivation of the Elasticity of Labor Demand

In order to derive an expression for the elasticity of labor demand in this model, I start by defining gross wages paid by the firm as $w_g \equiv (1 + \varphi)\, w$. Then, I write the elasticity of labor demand with respect to the *average* labor cost as:

$$e_{ALC} = -\frac{\Delta l/l}{\Delta w_g/w_g} = -\frac{\Delta l/l}{\Delta \varphi/(1+\varphi)} \tag{3.4}$$

This elasticity captures the percentage response of labor demand to a one percent change in labor costs. While this magnitude might be of interest in some particular cases, economic models usually define elasticities with respect to marginal changes in the price, because they give more general information.

The technical problem is that the labor cost functions has a discontinuity at the notch, so its derivative (the marginal cost of labor) is not defined at $l = l^*$. I approximate the "effective" marginal labor cost following the methodology proposed by Kleven and Waseem (2013). The idea is to consider what is the change in the marginal labor cost faced by the marginal buncher, relating the change in the tax rate, $\Delta \varphi$, with the change in labor demand, Δl, as follows:

$$\varphi^e = \frac{C\left(l^* + \Delta l^*\right) - C\left(l^*\right)}{\Delta l^*} \simeq \varphi_0 + \frac{\Delta \varphi}{\Delta l^*} l^* \tag{3.5}$$

This is depicted by the red dashed line in Figure 3.1, which links the two indifference points for firm B and has a slope equal to $(1 + \varphi^e)\, w$. An intuitive interpretation of 80 is that the cost of an additional unit of labor for firm B is the original marginal cost φ_0 plus the extra labor cost $\Delta \varphi$ that now has to be paid for *all* the existing employees (l^*). The latter term is averaged over the actual increase in employees, Δl^*. The size of this second term depends not only on the magnitude of the change in the labor cost, but also on the shape of the production function and the resulting isoprofit curves.

Given the above expression for φ^e, the elasticity of labor demand with respect to the marginal labor cost, evaluated at employment level l^*, can be written as:

$$e_{MLC} = -\frac{\Delta l^*/l^*}{\Delta \varphi^e/(1+\varphi^e)} \simeq -\frac{\left(\Delta l^*/l^*\right)^2}{\Delta \varphi/(1+\varphi)} \tag{3.6}$$

Notice that, in the final expression of e_{MLC}, the denominator is again the *average* cost of labor, while the numerator is squared.

3.2.4   Elasticity Estimation

Recall that, in equilibrium, $l_i \sim h(l)$. The small notch generated by an increase of $\Delta \varphi$ in the labor tax leads a group of firms to bunch. The marginal buncher now chooses to hire l^* employees, but it used to choose $l^{MB}\left(\theta^{MB}, \varphi^0\right) = l^* + \Delta l^*$, where Δl^* is related to $d\varphi$ through the shape of the density function $g(\theta)$. The term Δl^* is the length of the interval where no firms locate in Figure (3.2). Then, let the number of bunching firms be defined as

$$B = \int_{l^*}^{l^* + \Delta l^*} h(l)\, dl \approx h(l^*)\, \Delta l^*, \tag{3.7}$$

where the approximation relies on $\Delta\varphi > 0$ being small. Substituting (3.7) into (3.6), I obtain:

$$e_{MLC} = -\frac{(B/\left[h\left(l^*\right)l^*\right])^2}{\Delta\varphi/\left(1+\varphi\right)} \tag{3.8}$$

All the terms in expression (3.8) are either parameters or can be estimated empirically. The parameters are the value of the notch, $l^* = 50$ in this setting, and the change in the tax rate, $\Delta\varphi/\left(1+\varphi\right)$. To estimate the number of bunching firms, B, I construct a counterfactual density distribution that in turn provides the value of $h\left(l^*\right)$. Hence, the number of bunching firms is a sufficient statistic to estimate the elasticity of labor demand.

In order to construct the counterfactual density, I follow the standard methods from the bunching literature (e.g., Saez, 2010; Chetty et al., 2011; Kleven and Waseem, 2013). I first run a polynomial fit of the count of firms at each employee level (that is, the height of the bars in the histogram) on the number of employees:

$$S_i = \sum_{j=1}^{p} \beta_j \left(l_i - l^*\right)^j + \sum_{j=l_{lb}}^{l_{ub}} \gamma_j^0 \cdot \mathbf{1}\left\{l_i = j\right\} + \varepsilon_i^0 \tag{3.9}$$

where S_i is the number of firms with exactly i employees, $(l_i - l^*)$ is the distance of i from the threshold (so the notch is at 0), p is the order of the polynomial and l_{lb} and l_{ub} are the lower and upper bound of the excluded interval where the counterfactual diverges from the actual density. The counterfactual density is obtained by taking predicted values from the estimates of (3.9), excluding the γ_k shifters:

$$\widehat{S}_i = \sum_{j=0}^{p} \widehat{\beta}_j \cdot \left(l_i - l^*\right)^j \tag{3.10}$$

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):

$$\widehat{B} = \sum_{l=l_{lb}}^{l^*} \left|S_j - \widehat{S}_j\right| \qquad\qquad \widehat{H} = \sum_{l=l^*}^{l_{ub}} \left|S_j - \widehat{S}_j\right| \tag{3.11}$$

Determining the lower and upper bounds of the excluded region in a consistent way is critical for this estimation method to provide credible estimates. I follow the approach 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. I start by setting an arbitrary lower bound, l_{lb}, and then run equation (3.9) multiple times. Regarding the upper bound, in the first iteration I set $l_{ub} \approx l^*$, 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 l_{ub} by one employee and run equation (3.9) again as long as $\widehat{B} > \widehat{H}$. The process continues until it reaches a value of l_{ub} such that $\widehat{B} = \widehat{H}$. This procedure provides an estimate of the number of bunching (and missing) firms following the formulas in (3.11).

Since this estimation procedure is applied to the universe of French firms rather than a random sample, there is no sampling error and therefore I cannot construct the usual

confidence intervals around the estimates of B and H. To generate standard errors, I sample the residuals from regression (3.9) a large number of times (with replacement) to obtain bootstrapped standard errors. Finally, I apply the delta method to calculate standard errors around the elasticity estimates.

Optimization Frictions

Contrary to the prediction of the stylized model (pictured in Figure 3.2), there may be no "hole" in the firm size distribution above the notch. This would suggest that some firms are not able to adjust their number of employees as easily as others, and end up just above the cutoff despite the additional costs.

Optimization frictions have been a widely discussed issue in the bunching literature, as described in Chapter 1. The number of employees differs from reported revenue on several characteristics. First, it is a discrete variable, rather than continuous. Second, it is a stock instead of a flow. This matters dynamically because hiring a new worker today might have an impact in next year's number of employees, which is not the case for revenue. More generally, labor market rigidities make hiring and firing decisions less reversible than they would be in a frictionless world. Given all these reasons, it is important to allowing for the possibility that some firms may not be able to respond to the notch's incentives even though it would be optimal for them to do so.

Other sources of optimization frictions might come from searching and training costs that prevent firms from growing or laying off workers. Moreover, there might be uncertainty about the future so that there is a "region of inaction" in which firms do not adjust their factor demands in response to small shocks, as argued by Bloom (2009). In this particular context, a plausible hypothesis is that, once a firm has already been subject to the additional regulations, it cannot revert to the old situation by simply firing some employees and going back below the cutoff. For example, the committees that are created when crossing the 50-employee mark cannot be dissolved without approval from the unions. Thus, having been above the threshold in previous years could be a good predictor of locating in the interval just above 50 employees.

I incorporate the possibility of optimization frictions into the elasticity estimation following the approach proposed by Kleven and Waseem (2013). I define α as the proportion of firms locating in the interval $(l^*, l_{ub}]$, compared to the counterfactual density. I use this measure to re-weight the estimates so that $\widehat{B}_F = \frac{\widehat{B}}{1-\alpha}$, where the subscript F indicates that the new estimator accounts for optimization frictions. Then, I introduce $\widehat{B}_F$ into the elasticity formula:

$$e_F = -\frac{\left(B_F / \left[h\left(l^*\right)l^*\right]\right)^2}{\Delta\varphi / \left(1+\varphi\right)} \tag{3.12}$$

The estimates of e_F can be interpreted as an upper bound of the structural elasticity parameter.

3.3 Institutional Context and Data

3.3.1 Labor Regulations in France

There are at least 34 different labor regulations in France that only apply to firms with more than 50 employees (Attali, 2008). Some other regulations apply only to firms above 10 or 20 employees. Comprehensive summaries of all these regulations are presented in Ceci-Renaud and Chevalier (2010) and in the online appendix of Garicano, LeLarge and van Reenen (2013). In this paper, I focus on the 50-employee threshold, which is the most important in terms of the difference in average labor costs on each side of the cutoff. I summarize below the key regulations that apply above this threshold.

Firms with more than 50 employees have to set up a Works Council (*"comité d'entreprise"*), which meets bi-monthly to discuss workplace issues. This committee is comprised of 3-4 representatives of the employees, usually appointed by the unions, who get 20 paid hours per month to perform committee duties. Firms also have to set up committees on hygiene and safety of working conditions. Additionally, they have to establish a profit-sharing scheme, by which employees receive a pre-determined share of annual profits (which varies by sector). Another law regulates collective dismissals, defined as the separation of 9 or more employees in any given month. A firm that wants to carry out a collective dismissal needs to design a "social plan", monitored by the Ministry of Labor and negotiated with the unions, in order to facilitate re-employment of laid-off workers. There is evidence that this process usually takes up substantial administrative resources from firms.

It is hard to provide a precise measure of the costs that these regulations impose on firms, in part because there is probably a great deal of heterogeneity. In order to obtain elasticity estimates using the methodology outlined above, it is necessary to assign a specific labor tax rate to this package of regulations. A reasonable source for this estimate is Attali (2008), a report commissioned by the French government to implement an ambitious reform program. Attali (2008) estimates, based on a large survey of firm managers, that the package of regulations that kicks in at 50 employees implies an extra labor cost equivalent to 4% of the total wage bill for the average firm in the 50-250 employees range.

3.3.2 Data

I use data from Amadeus, a comprehensive database of European businesses put together by Bureau van Dijk, a market research company (www.bvdinfo.com), already described in Chapter 1. The dataset covers annual accounting reports from French firms for the period 2002-2009. The information available for each firm in each year includes: business 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.

Table 3.1 reports the number of firms reporting their number of employees in each year during this period. The number of firms grows from just above 300,000 in 2002 to 900,000 in 2008 and almost 1.8 million in 2009. Part of the trend is probably due to actual growth in business registration, but the bulk of the change is due to other factors. First, the online version of Amadeus accessed for this paper (through Wharton Research Data Services) only maintains firms in the dataset for 4 years after their last report. Thus, if a firm disappeared

in 2008, it would not be in the dataset in 2013. Second, in some cases Amadeus does not digitalize the entire financial statements when firms submit them in paper form, but this is not an issue when the submission is electronic. A change in the proportion of firms submitting electronically is likely to explain the doubling in the number of firms from 2008 to 2009. As can be seen in Table 3.1, most of the increase in 2009 is due to firms with 10 or fewer employees. In any case, to avoid any sample selection issues, when I pool data for several years in the empirical analysis I use the period 2002-2008.

3.4   Results

3.4.1   Graphical Evidence

I begin by studying the cross-sectional distribution of firm size, measured by the number of employees. In the absence of size-contingent labor regulations, the firm size distribution (FSD) should follow a smoothly declining distribution, approximately lognormal (Cabral and Mata, 2003). Hence, any deviations of the FSD from a smooth counterfactual distribution can be attributed to the existence of the regulatory cutoff.

Figure 3.3 shows the firm size distribution around the 50-employee threshold using pooled data from the years 2002-2008. Pooling increases sample size and it is consistent with the institutional context because the same set of labor regulations was applicable throughout the period. The year-by-year distributions look very similar, as shown in Figure 3.7. There is a large spike in the density starting at 49 employees, and then a drop from 50 onwards.[9] The number of firms with 49 employees more than doubles the number of firms with 50 employees. The bunching response indicates a behavioral response to labor regulations, as predicted by the theoretical framework. Moreover, the lack of a "hole" in the density above the notch suggests that there are non-negligible optimization frictions that prevent some firms from adjusting. In the next subsection, I quantify the bunching response and provide estimates for the implied elasticity of labor demand both under the assumptions of optimization frictions and the stylized frictionless model. Figure 3.4 shows the firm size distribution around the two other regulation thresholds at 10 and 20 employees. There seems to be bunching in both cases, although the jump in the density at the notch is proportionally smaller.

The two graphs in Figure 3.5 describe the dynamics of firm growth in France. The top panel plots the proportion of firms that grow in the following year (vertical axis) against their current size (horizontal axis), defining growth as having more employees in year $t + 1$ than in year t. The proportion increases with the number of employees, but it drops sharply below each of the three regulatory thresholds. In particular, the proportion of growing businesses drops from 35% for 40-employee firms to 20% for 49-employee firms. The dashed lines indicate 95% confidence intervals for each bin, which indicate that the downward jump is statistically significant. In the bottom panel, the y-axis variable is the proportion of firms that keep the same number of employees in year $t + 1$ as they had in year t. The proportion decreases with firm size, but it jumps up from 15% for 40-employee firms to about 30% for

---

[9]It seems that the additional labor regulations applied to firms with "50 and more" employees for the period 2002-2008. The data for 2009 actually shows the largest spike at 50 employees. I am not certain of whether this is due to a change in one specific regulation or a larger subset.

49-employee firms (also statistically significant).

To sum up, there is strong evidence that some firms strategically reduce their growth when they reach 49 employees in order to avoid being subject by the additional labor regulations.

3.4.2 Elasticity Estimations

Using the method described in the empirical strategy section, I construct a counterfactual firm size distribution, shown in Figure 3.6. The blue connected dots represent a histogram with the same data as Figure 3.3. The orange dashed line depicts the counterfactual distribution. The vertical dashed lines indicate the lower and upper bounds of the excluded region, respectively. Based on visual observation, I pick $l_{lb} = 40$ and the estimation routine determines that $l_{ub} = 63$. The estimated number of bunching firms over the 9-year period is $\widehat{B} = 8,223$, from a sample of $208,058$ firms between 20 and 80 employees. To approximate the change in labor costs associated with the labor regulations, I use the estimates from the Attali (2008) report, such that $\Delta\varphi/\left(1+\varphi\right) = 4\%$.

Table 3.2 reports the elasticity estimates for the pooled 2002-2008 data and also for each year's sample. The "no frictions" estimate for the pooled data is $\widehat{e}_{NF} = 0.055$ with standard error 0.032, so it is statistically significant at the 10% level. The point estimate allowing for optimization frictions is $\widehat{e}_{NF} = 0.572$ with std. error 0.668, which is not statistically distinguishable from zero (nor one). The point estimates are remarkably similar for all the years between 2002 and 2009, with slightly larger variation on the standard errors. The "no frictions" elasticity is marginally significant in three years, while the "frictions" elasticity is not significant in any year.

In a thorough review on the topic, Hamermesh (1993) reviews the existing estimates of the labor demand elasticity from previous studies and concludes that the average is around 0.3, with a confidence $(0.15, 0.75)$. This puts the "no frictions" estimates below the lower bound of the distribution, which is somewhat surprising considering the sharp bunching observed in the size distribution. The key to understanding this apparent contradiction is that a notch creates very strong incentives, so with a labor demand elasticity of 0.3 we would have observed much stronger bunching at 49 employees.

3.4.3 Other Margins of Adjustment

It is clear that the stylized one-input model that we have imposed on the data might be driving these low estimates of the structural elasticity. A natural extension is to include other production inputs in the model, such as capital or materials, and study the elasticity of substitution around the notch.

As a preliminary inquiry into this possible adjustment margins, Figure 3.8 plots the distributions of fixed assets per employee (a proxy for the capital-labor ratio, k/l) and material input expenses per employee. The graphs show median values for groups of firms with the same number of employees for the pooled sample 2002-2008. The top panel of Figure 3.8 shows that the median capital-labor ratio increases with the number of employees, but there is a sharp drop exactly at the notch, going from a median of €15,000 in fixed assets per employee for 49-employee firms to a median of €12,000 for 50-employee firms. In the bottom

panel, there is also a small break in the pattern of median material inputs per employee at the notch, but it is less clear that the overall trends differ on either side of the threshold.

The positive correlation between fixed assets and material expenditures and locating just below the regulatory threshold suggests that bunching firms may be substituting away from labor and into other inputs as they approach the cutoff. In future work, I plan to derive the model with multiple production inputs and a constant elasticity of substitution (CES) technology. Hamermesh (1993) shows how to bring this model to the data in order to estimate both the labor demand elasticity and the elasticity of substitution between labor and other inputs. The latter elasticity is interesting in its own right, and it will also allow me to disentangle to what extent those margins of adjustment bias my estimates of the labor demand elasticity towards zero.

3.5   Welfare Implications

Using the estimates for the elasticity of labor demand from the previous section, I can make an estimation of the deadweight loss (DWL) associated with the additional labor market regulations. In particular, I attempt to estimate the welfare loss due to the fact that, in response to the higher labor costs, some firms reduce their demand for labor but workers do not fully capture the surplus lost by firms.

Recall that firms in the model are heterogeneous in their managerial productivity θ, so the standard DWL formulas for a representative agent cannot be applied. I then calculate the individual DWL for a each firm and then aggregate over all firms. Given that the model assumes an infinitely elastic labor supply curve, the DWL is simply the area below each firm's demand curve and above the market wage, over the interval $(l_i(\theta_i, \varphi^0), l_i(\theta_i, \varphi^1))$:

$$
\begin{aligned}
DWL_i &\simeq \frac{(\varphi_1 + \varphi_0)\,w}{2}\left[-\Delta l_i\right] \\
&\simeq \frac{(\varphi_1 + \varphi_0)\,w}{2}\left[-e\frac{\Delta\varphi}{(1+\varphi)}l_i(\theta_i)\right]
\end{aligned}
\tag{3.13}
$$

When aggregating for the entire economy, I do not include low productivity firms because they are not affected by the new regulations. The impact on bunching firms is larger than on high-productivity firms because they change their labor demand more, but the same formula applies to all firms with productivity in the range $\theta \in [\theta^*, \theta_{max})$. Then, aggregate DWL is given by:

$$
DWL = \int_{\theta^*}^{\theta_{max}} \frac{(\varphi_1 + \varphi_0)\,w}{2}\left[-e\frac{\Delta\varphi}{(1+\varphi)}l(\theta)\right]d\theta
\tag{3.14}
$$

This calculation yields an aggregate deadweight loss 0.42% of the total wage bill, when the "no frictions" elasticity estimation is considered, and of 1.78% when the upper bound estimate (allowing for frictions).

In a study that analyzes the same set of labor regulations, Garicano et al. (2013) structurally estimate a model based on Lucas (1978). They calculate that the welfare loss associated with the French labor market regulations is 0.8% of GDP when assuming flexible wages and up to 5.1% of GDP when assuming wage rigidities. In the flexible wage case (similar to

the assumptions made in this paper), the output loss is small and the drop in efficiency comes almost entirely from the tax aspect of regulations. (Garicano et al., 2013) point out that the tax could be considerecd a transfer to workers, but they provide suggestive evidence showing that workers do not place high value on the amenities obtained from more union power and new committees. In the rigid wage model, the regulations lead to higher unemployment and thus lower substantially lower output, which accounts for most of the efficiency loss in that case.

3.6 Conclusion and Future Work

I have studied the impact of labor regulation thresholds on the production behavior of French firms. About 34 different labor regulations apply only to firms with 50 or more employees, giving firms an incentive to bunch just below this threshold. Using a stylized model with only one factor of production, I have derived an elasticity of labor demand that can be estimated directly using the number of bunching firms as a sufficient statistic.

I obtain a point estimate of $e_{NF} = 0.055$ for the elasticity of labor demand in France in the period 2002-2008, which is statistically different from zero at the 10% level. Making an adjustment for the possibility that some firms do not respond to the regulations due to optimization frictions, I obtain a point estimate of $e_F = 0.572$, which could be interpreted as an upper bound for the long-term structural elasticity, although it is imprecisely estimated (the standard error is 0.668).

These point estimates are considerably below other labor demand elasticities estimated in the literature, which according to Hamermesh (1993) tend to be in the interval $(0.15, 0.75)$. An intuitive explanation for these low estimates is that bunching firms might be adjusting their production by increasing the use of other inputs instead of labor. As an example, I have shown that median fixed assets per employee are much higher for bunching firms than for firms immediately above the threshold.

The next step in this research project is to explore potential extensions to the model to make it more realistic and allow a better identification of the structural elasticities. One extension discussed in the main text is to consider multiple inputs in the production function, and study the elasticity of substitution between labor and other inputs. Another possible extension is to endogenize wage determination by allowing the labor supply curve to be upward sloping. In a model with endogenously determined wages, firms above the threshold would pay lower wages to their employees in equilibrium, because the regulations act as a benefit that the employees receive(Summers, 1989). A countervailing effect is that, on average, larger firms tend to pay higher wages. This relationship is likely to be monotonic over firm size, while the "benefit" effect would create a jump at the threshold. Finding a way to test this hypothesis using the incentives created by the regulatory cutoff seems to be a promising avenue for future research.

A more ambitious extension of the model would be to add dynamics and uncertainty. This would allow me to incorporate the possibility of good versus bad economic conditions, which complicates the hiring and firing decisions of the firm as expectations come into play. In this dynamic model, firm owners would need to trade-off present versus future benefits and costs.

Tables

Table 3.1: Amadeus Data, French Firms

	0-10 Employees	11-100 Employees	100+ Employees	Total
2002	260,084	38,559	6,572	305,215
2003	449,474	61,639	8,450	519,563
2004	506,774	68,298	8,920	583,992
2005	566,252	70,202	9,141	645,595
2006	646,454	62,486	8,925	717,865
2007	739,836	61,299	9,109	801,244
2008	827,869	58,484	8,408	894,761
2009	1,702,772	67,455	8,606	1,778,833

Note: this table shows the evolution of the sample size in the Amadeus dataset for French firms. Only firms with a nonmissing value for the number of employees are counted. A change in the proportion of firms submitting electronically is likely to explain the doubling in the number of firms from 2008 to 2009. As can be seen in the table, most of the increase in 2009 is due to firms with 10 or fewer employees. To avoid any sample selection issues, when I pool data for several years in the empirical analysis I always use the period 2002-2008.

Table 3.2: Labor Demand Elasticity Estimates

	e_{NF}	e_F	B	H	l_{lb}	l_{ub}	N
Pooled Data							
2002-2008	0.055*	0.572	8,223.5	8,530.1	40	63	208,058
	(0.032)	(0.668)					
Annual Data							
2002	0.050	0.564	784.1	900.1	40	64	20,439
	(0.040)	(3.759)					
2003	0.065*	0.565	1,289.3	1,363.2	40	63	30,801
	(0.037)	(0.622)					
2004	0.057*	0.506	1,318.3	1,257.0	40	61	33,480
	(0.031)	(0.516)					
2005	0.055	0.508	1,323.0	1,421.5	40	63	33,921
	(0.045)	(0.995)					
2006	0.062*	0.744	1,282.4	1,304.0	40	65	30,690
	(0.034)	(1.737)					
2007	0.041	0.424	1,066.0	1,293.8	40	63	30,318
	(0.027)	(0.534)					
2008	0.056	0.677	1,140.2	1,094.5	40	63	28,409
	(0.043)	(3.066)					

Note: e_{NF} and e_F are the labor demand elasticities assuming no frictions and frictions, respectively. B is the number of firms above the counterfactual density in the range $l \in (l_{lb}, l^*)$, where l is number of employees, l_{lb} is the lower bound of the excluded region (used to construct the counterfactual) and l^* is the regulatory threshold of 50 employees. H is the missing number of firms below the counterfactual density in the range $l \in (l^*, l_{ub})$, where l_{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 $l \in (20, 80)$ in each year. Significance levels: *** = 1%, ** = 5%, and * = 10%.
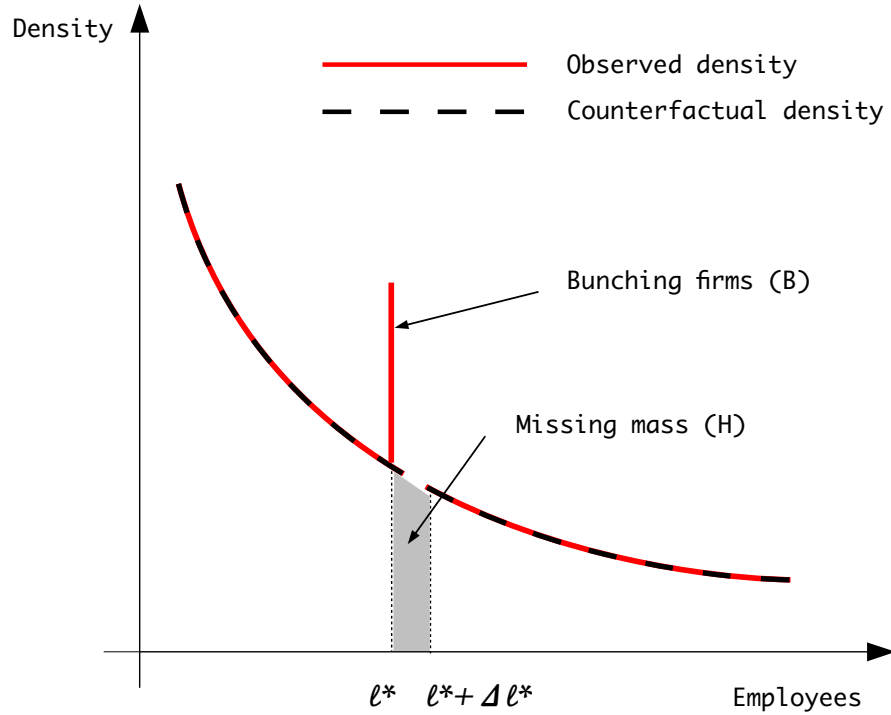
Figures

Figure 3.1: Discontinuity in Labor Costs
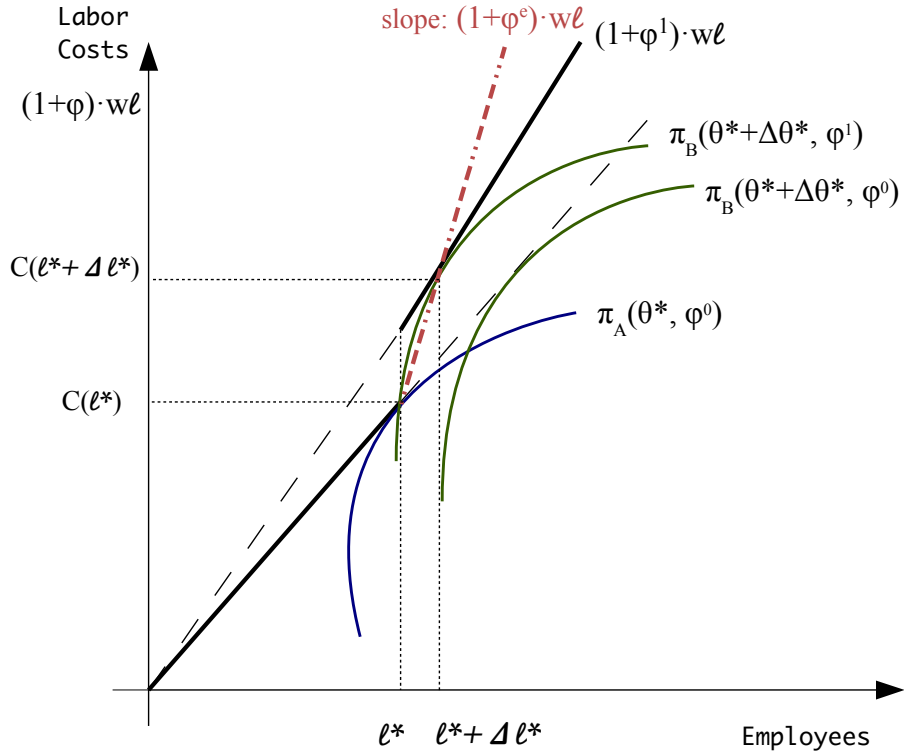
Note: this diagram shows equilibrium in a model with a notch in labor costs at l^* employees. The vertical axis shows total labor costs: wage (w) times employees (l) multiplied by one plus the tax on labor $(1+\varphi)$. The horizontal axis plots the number of employees. The rays from the origin depict total labor costs under low and high tax rates, φ^0 and φ^1. The higher tax only applies to firms with more than l^* employees, so the schedule of labor costs is determined by the solid black lines, with a discontinuous upward jump at l^*. The curves represent isoprofit curves for firm A and B, which have managerial productivity levels $\theta_A = \theta^*$ and $\theta_B = \theta^* + d\theta^*$, respectively.

Figure 3.2: Theoretical Employees Distribution with a Notch

Note: this figure depicts the theoretical revenue distribution before and after the introduction of labor regulations. In the standard scenario, the distribution of number of employees is smoothly decreasing, as depicted by the dashed (black) line. When additional labor regulations are introduced, firms with more than l^* employees face higher average labor costs. A group of firms in an interval above l^* respond to the new regulations by reducing their size to have exactly l^* employees. 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 Δl^* above the threshold.

Figure 3.3: Firm Size Distribution
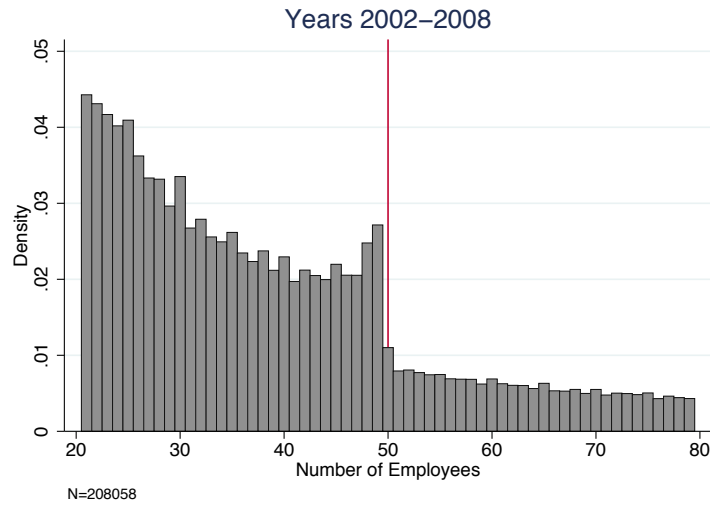
Note: this histogram shows the distribution of firm size in France, measured by the number of employees. The vertical red line indicates that firms above that size (50 employees) are subject to a set of 34 labor regulations. The graph pools data for the period 2002-2008, throughout which the regulations were applicable. Bin width is one, meaning that the histograms show the raw data without any adjustments.

Figure 3.4: Bunching at Other Thresholds
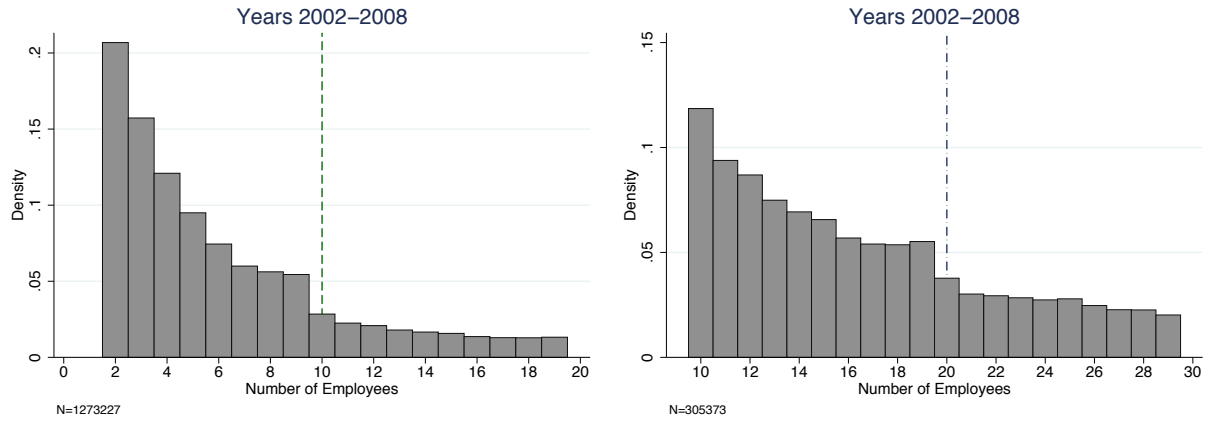
Note: these two histograms show the distribution of firm size in France, measured by the number of employees. The vertical dashed lines indicate that firms with more than 10 employees (left) or 20 employees (right) are subject to additional labor regulations. The graphs pool data for the period 2002-2008, throughout which the regulations were applicable. Bin width is one, meaning that the histograms show the raw data without any adjustments.

Figure 3.5: Employment Growth Patterns and Persistence

Proportion of Firms Growing the Following Year
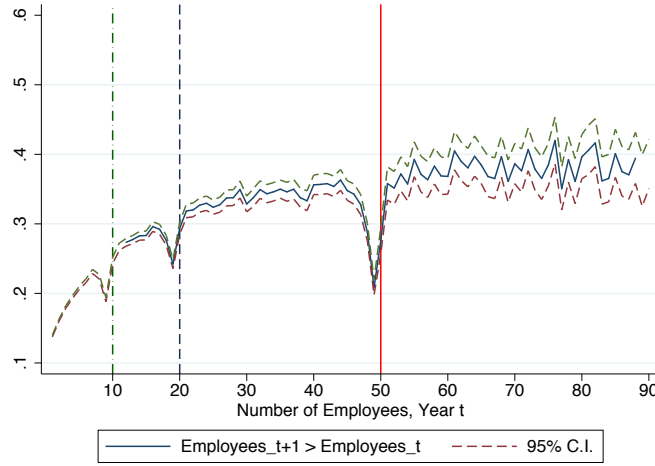
Proportion of Firms Keeping the Same Number of Employees
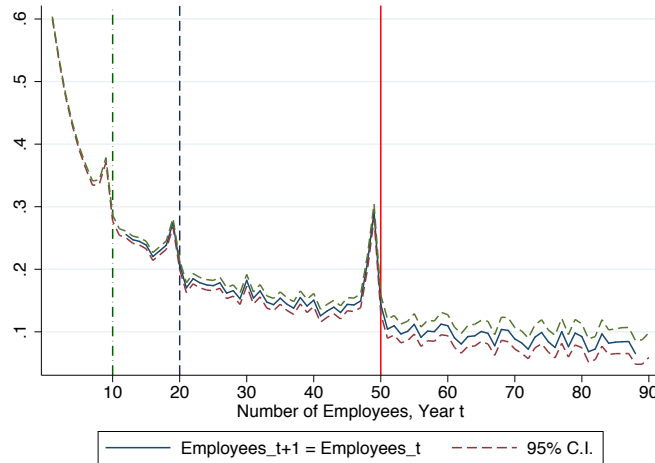
Note: the top panel shows bin averages for the dummy variable $grow_{it} = 1\{l_{i,t+1} > l_{it}\}$, where l_{it} is the number of employees of firm i in year t. The bottom panel shows bin averages for the dummy variable $stay_{it} = 1\{l_{i,t+1} = l_{it}\}$. The dashed curves represent 95% confidence intervals around each bin average. The green vertical dash-dot line indicates the 10-employee threshold at which some labor regulations kick in. Similarly, the blue dashed line indicates the 20-employee threshold at which additional regulations apply, and the solid red line indicates the 50-employee threshold at which another 34 regulations apply.

Figure 3.6: Counterfactual Firm Size Distribution, 2002-2008
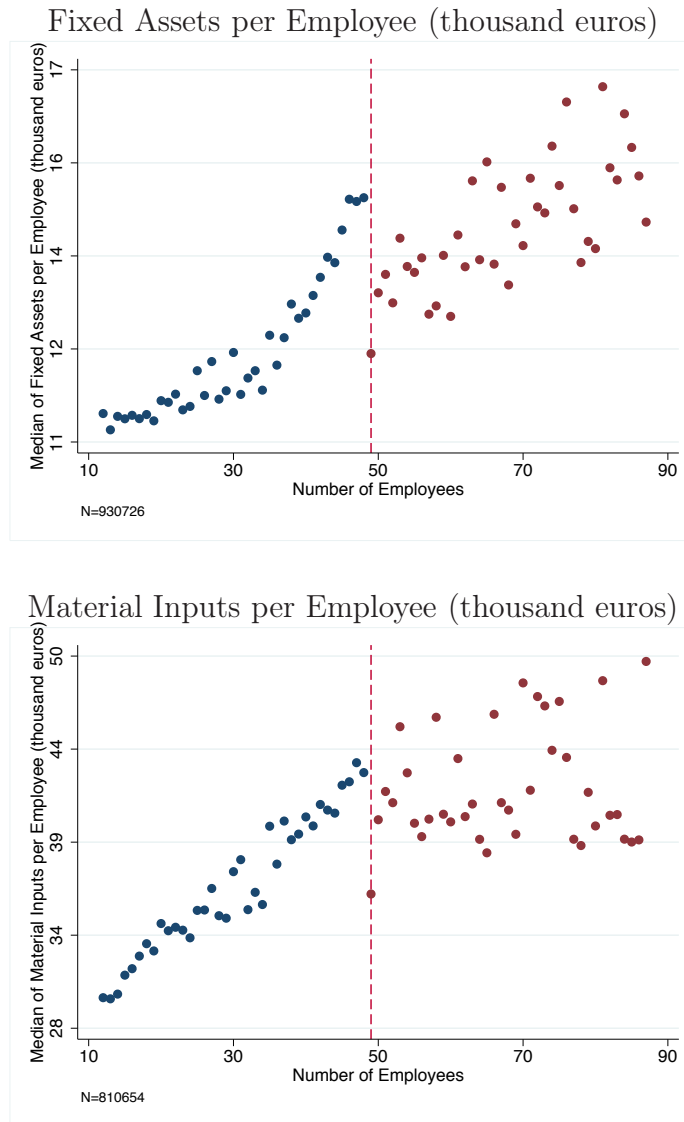
Years 2002-2008

N=208058
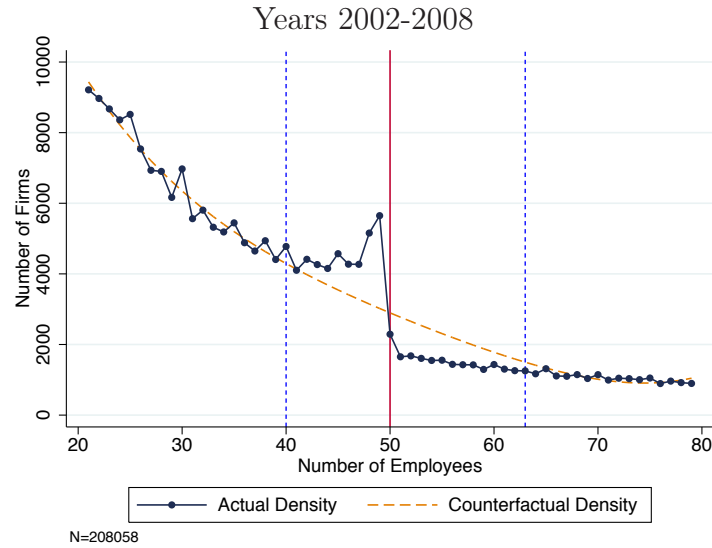
Note: the dots connected by a solid blue line represent a histogram of the number of employees, using exactly the same data as Figure 3.3. The orange dashed line shows the counterfactual firm size distribution. The vertical dotted blue lines indicate the lower and upper bounds of the excluded region, l_{lb} and l_{ub}. To determine these values, I first fix the value of $l_{lb} = 40$ employees and then fit a polynomial fit to the density, starting with $l_{ub} \approx l^* = 50$. I then increase the value of l_{ub} in unit steps until the are between the observed density and the counterfactual in the range $(l_{lb}, 50)$ equals the are between the two densities in the range $(50, l_{ub})$.

Figure 3.7: Counterfactual Firm Size Distribution, Year by Year
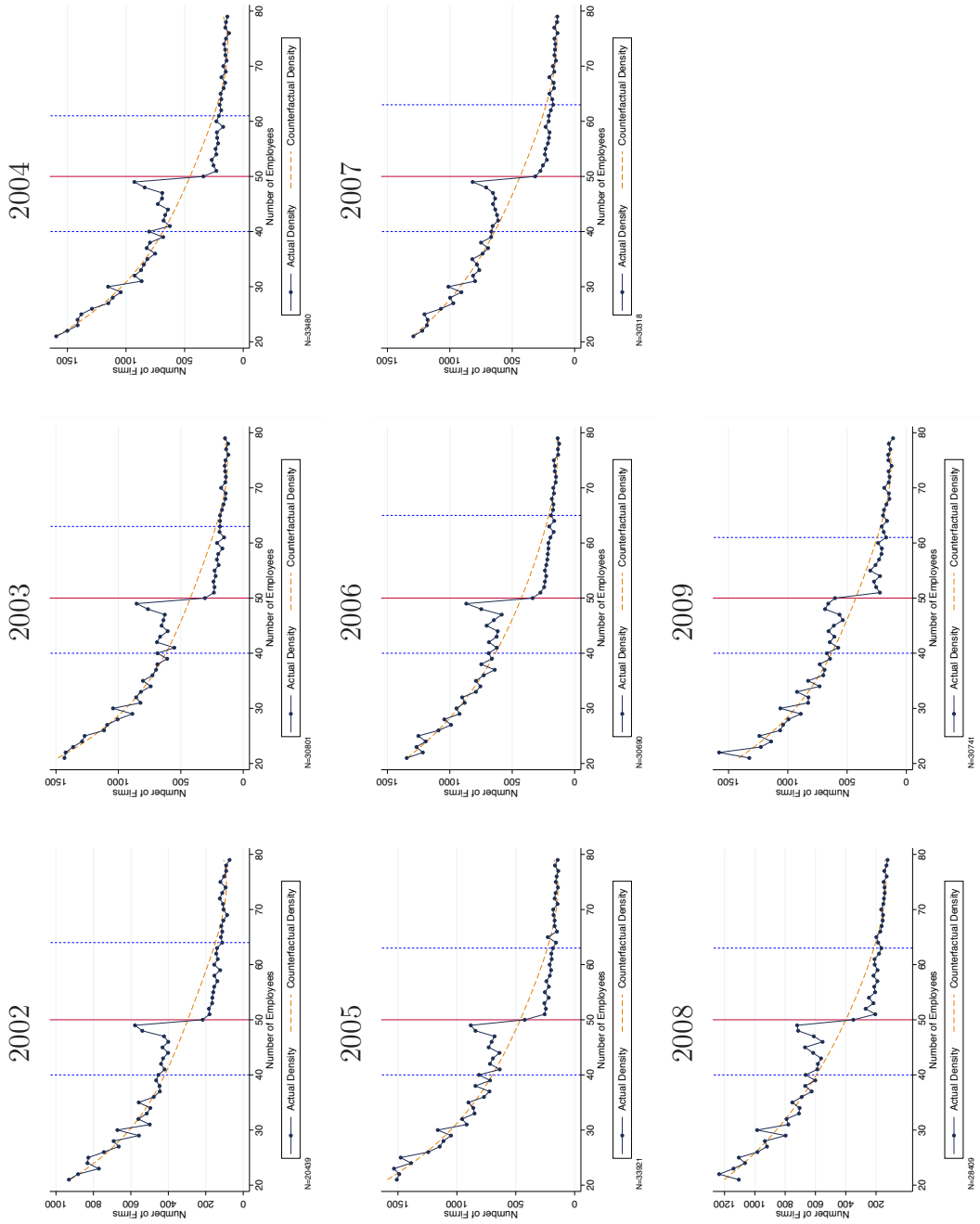
Note: these annual plots are constructed in the same way as Figure 3.6.

96

Figure 3.8: Use of Other Production Inputs

Fixed Assets per Employee (thousand euros)

Material Inputs per Employee (thousand euros)

Note: the top panel shows median fixed assets per employee (in thousand euros) and the bottom panel shows median material inputs per employee (in thousand euros). The bin width to calculate the medians is one employee. The vertical red dashed line indicated the 50-employee threshold above which labor regulations are stricter.

Bibliography

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.

Andreoni, James, Brian Erard, and Jonathan Feinstein, "Tax Compliance," *Journal of Economic Literature*, June 1998, *36*, 818–860.

Attali, Jacques, "Rapport de la Commission pour la libération de la croissance française ("Report for the Commission for the Liberation of French Growth")," 2008. Report commissioned by French President Nicolas Sarkozy.

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.

Bloom, Nicholas, "The impact of uncertainty shocks," *Econometrica*, 2009, *77*, 623–685.

BOE, "Ley de Presupuestos Generales del Estado," Several years. Boletín Oficial del Estado.

Braguinsky, Serguey, Lee Branstetter, and Andre Regateiro, "The Incredible Shrinking Portuguese Firm," July 2011. NBER Working Paper 17265.

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.

Ceci-Renaud, Nila and Paul-Antoine Chevalier, "L'impact del seuils de 10, 20 et 50 salariés sur la taille des entreprises françaises," *Économie et Statistique*, 2010, *437*, 29–45.

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.

_ , "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Analysis," *Annual Review of Economics*, June 2009, *1*, 451–488.

____ , "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.

Garicano, Luis, Claire LeLarge, and John van Reenen, "Firm Size Distrotions and the Productivity Distribution: Evidence from France," February 2013. NBER Working Paper 18841.

Guner, Nezih, Gustavo Ventura, and Yi Xu, "Macroeconomic Implications of Size-Dependent Policies," *Review of Economic Dynamics*, 2008, *11*, 721–744.

Hamermesh, Daniel, "The demand for labor in the long run," *Handbook of Labor Economics*, 1986, *1*.

____ , *The Demand for labor*, Princeton, NJ: Princeton University Press, 1993.

Heckman, James and Carmen Pages, "Law and Employment: Lessons from Latin America and the Caribbean," December 2003. NBER Working Paper 10129.

Hsieh, Chang-Tai and Peter Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, November 2009, *CXXIV* (4), 1403–1448.

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.

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.

LaCaixa, "Por Qué No Aumenta la Dimensión de la Empresa Española?," May 2012. La Caixa Research Department.

Lee, David and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, June 2010, *48*, 281–355.

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.

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.

Schivardi, Fabiano and Roberto Torrini, ""Identifying the effects of firing restrictions through size-contingent differences in regulation," *Labour Economics*, 2008, *15*, 482–511.

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.

Summers, Lawrence, "Simple Economics of Mandated Benefits," *American Economic Review, Papers and Proceedings*, May 1989, *79* (2), 177–183.