WP/18/179

IMF Working Paper

Size Dependent Policies, Informality and Misallocation

by Era Dabla-Norris, Laura Jaramillo Mayor, Frederico Lima, and Alexandre Sollaci

INTERNATIONAL MONETARY FUND

 WP/18/179

IMF Working Paper

Fiscal Affairs Department

Size Dependent Policies, Informality and Misallocation

Prepared by Era Dabla-Norris, Laura Jaramillo Mayor, Frederico Lima, and Alexandre Sollaci*

Authorized for distribution by Era Dabla-Norris

August 2018

Abstract

We examine the effect of size-dependent policies in developing economies by focusing on a set of regulations that are applicable to firms with 20 or more formal employees in Peru. Firms can adjust to the regulations by (a) reducing their size, (b) shifting employment composition, or (c) splitting into subunits that fall below the regulatory threshold. We show that these actions are consistent with observed discontinuities in the distributions of firm size and employment composition. We extend the framework proposed by Garicano et al. (2016) to model and estimate the Peruvian economy and perform counterfactual exercises. Size-dependent regulations are costly for the economy, especially in the presence of labor market rigidities, and lead to lower aggregate wages, profits, and output. We also find that access to informal labor does not mitigate the economic impact of the size-dependent regulations, as the increase in informal employment is largely offset by a decline in formal employment.

JEL Classification Numbers: E24, E26, J46

Keywords: size-dependent policies, misallocation, productivity, informality

Author's E-Mail Address: edablanorris@imf.org; ljaramillomayor@imf.org (corresponding author); flima@imf.org; asollaci@uchicago.edu

* We thank Rodrigo Azuero, Claudia Berg, Mariano Bosch, Shafik Hebous, Suhaib Kebhai, Alexander Klemm, Daniel Garcia-Macia, Chris Papageorgiou, Luca Ricci, Marta Ruiz, Alasdair Scott, Jesica Torres, Frederick Toscani, and participants in seminars at the IMF, IADB, and the University of Chicago for helpful comments and discussions. We thank the IADB and SUNAT for help in accessing administrative firm data. Pooja Karnane provided excellent research assistance.

Contents

List of Figures

List of Tables

1 Introduction

Size dependent firm policies — policies that only affect firms above a given size threshold — are a common type of legislation in many countries. Although they are often intended to spare smaller firms from onerous regulation, such policies involve trade-offs that need to be taken into consideration. Indeed, size-dependent policies can create incentives for firms to remain sub-optimally small, potentially misallocating labor across firms and impacting aggregate productivity. Importantly, the macroeconomic effects of such regulations depend on other structural characteristics in the economy. In the case of developing countries, the presence of informality and weak government enforcement capacity can allow firms to circumvent regulations, while labor market rigidities may amplify their negative economic impact.

Several recent papers find that size dependent policies in advanced countries impose a significant cost to the economy by misallocating labor to smaller, less productive firms. For example, Garicano et al. (2016) explore the effect of stringent labor regulations affecting firms with more than 50 workers in France, and show that these regulatory kinks have a large impact on aggregate output and employment, especially if wages are sticky.[1] However, the implications of this work for developing economies are unclear, as firms in these countries have other ways of circumventing regulations, potentially mitigating their impact. For example, Bertrand et al. (2017) examine how large Indian firms use contract workers to get around stringent labor protection regulations, and argue that access to contract workers increases firm employment and investment.[2]

Our main goal in this paper is to measure the aggregate effect of size-dependent policies when firms have access to these additional margins of adjustment. We focus our attention on Peru, which is an ideal setting to study these interactions for several reasons. First, Peru has very stark size dependent regulations. Firms with more than 20 formal employees are required to comply with a number of extra regulations that do not apply to smaller firms, including a mandatory profit sharing requirement. These regulations are costly for firms, and there is a strong incentive to avoid them. Second, there is a large degree of labor informality in Peru, and the government is not able to enforce regulations evenly across all firms (Viollaz, 2018).

[1] Gourio and Roys (2014) examine the same firm size distortion, and use indirect inference to estimate a structural model that incorporates a sunk cost of complying with regulations. See also Schivardi and Torrini (2008) and Onji (2009) for examples from Italy and Japan, respectively.

[2] Amirapu and Gechter (2017) also study the impact of size-dependent labor regulations on firm size in India, and find evidence that regulatory costs are associated with higher corruption.

This allows Peruvian firms several margins of adjustment to bypass the regulations. Finally, unlike other countries, there exists survey data on firm employment that captures some types of informal workers, allowing us to measure directly how firms change labor composition to adjust to the regulatory threshold.

In this paper, we study regulations that only apply to firms with more than 20 formal, salaried employees in Peru. Firms circumvent these regulations through three avenues: (a) by reducing employment of formal workers to stay below the threshold; (b) by substituting towards informal labor, which does not count towards the threshold; and (c) by splitting into subunits that fall below the regulatory threshold. To model these decisions, we extend the framework proposed by Garicano et al. (2016) to include both formal and informal employment, and to allow firms to split into smaller subunits to avoid incurring regulatory costs, in a setting with and without labor market rigidities. We estimate this structural model to fit the observed firm size distribution in the Peru, and then use it to perform counterfactual exercises.

We model informality in the same spirit as Levy (2008), who distinguishes between three types of workers: (a) *formal, salaried workers* that have a formal employment contract, which makes them easier to monitor and control; (b) *informal, non-salaried workers* that have a valid contract of a different nature, often short-term (e.g., a traineeship or a service provision contract), which makes them more difficult to monitor and control, and can therefore make them less productive; and (c) *informal, illegal workers* that do not have a contract, are not declared by the firm to tax or labor authorities, and are effectively "off-the-books".[3]

Our results show that the size dependent regulations, as currently designed, are costly for the Peruvian economy. They hurt the economy by creating incentives for talented managers to hire sub-optimally (hiring fewer workers, and hiring less productive non-salaried workers), thereby reducing output, while at the same time boosting the number of smaller, less productive firms. The model shows that introducing these regulations is associated with a 0.4 to 1 percent decline in aggregate wages, and a 3 to 4 percent decline in aggregate profits. While the impact on employment and output remains small provided wages are fully flexible, we find that the regulations reduce output by up to 1 percent of GDP if we match the observed degree of wage rigidity in Peru. Our results show that the negative effects on output would be considerably worse if the threshold was lowered, and the policy applied to a larger share of firms. Overall,

[3]Although firm informality is also prevalent in the region (e.g., Ulyssea (2018) in the case of Brazil), we do not focus on informal firms since they are not captured in our data. However, informal firms tend to have fewer than 20 workers, and therefore are unlikely to be affected by the size-dependent regulations we study.

smaller firms win from the policy, while larger firms and workers lose. Smaller firms with fewer than 20 workers benefit from a lower overall wage in the economy, and thus are able to expand employment and increase profitability. At the same time, medium-sized firms that are close to threshold — with between 20 and 35 workers — are significantly worse off because they see a large decline in employment and a shift in employment composition towards less productive workers, which curtails output. Larger firms are also negatively affected, especially through a severe reduction in profitability. Finally, workers are worse off because they face lower wages and higher unemployment.

We also examine how the size-dependent regulations interact with labor informality. We find that introducing labor informality as an additional margin of adjustment does not change the aggregate impact of the regulations, as it leads to offsetting effects on labor demand. On one hand, labor demand increases since some medium-sized firms that are affected by the regulations opt to expand production by employing informal workers. On the other hand, labor demand decreases because the more productive of these medium-sized firms hire fewer formal workers. These two effects cancel each other out, suggesting that the presence of informality does not affect the cost of the regulations.

These results stand in contrast with other studies of size-dependent regulations in developing countries, which often do not find a significant impact of regulations on firm behavior (e.g., Hsieh and Olken, 2014). A potential explanation is that the interest in enforcing the regulations lies with the workers themselves in the case of Peru. In particular, the profit sharing requirement is widely reported in the media, and is enshrined in the Constitution. In contrast, many of the other policies previously examined in the literature are primarily enforced by the government, where capacity constraints often limit enforcement actions to larger firms.

The rest of the paper proceeds as follows. Section 2 presents the size-dependent regulations, and section 3 discusses the empirical evidence on how these affect firm decisions. Sections 4 and 5 present the baseline model, and the extension to firm splitting, respectively, and section 6 discusses the estimation steps. Section 7 shows the estimated counterfactuals, and section 8 concludes.

2 Institutional setting

Mandatory profit sharing has a long tradition in Peru, and has often been justified as a way to increase worker productivity and improve income redistribution. The current legal basis for this requirement is enshrined in article 29 of the Constitution of 1993, which sets out a general right of workers to participate in firm profits, and in legal decree 892 of 8 November 1996, which regulates how mandatory profit sharing should take place.

The current law requires that firms with positive before-tax profits distribute some of these profits to their employees in accordance with the number of days they worked and their base salary. Firms with fewer than 21 employees are exempt, and the overall share of profits to be distributed varies by industry, ranging from 10 percent for manufacturing, fishing and telecommunications, 8 percent for mining, retail, wholesale and restaurants, and 5 percent for other industries.[4,5] Firms can deduct these distributed profits from taxable income. For workers, distributed profits are subject to personal income taxation, but are exempt from social security contributions, unlike their wage compensation.[6]

Importantly, for the purpose of determining whether a firm is exempt from this requirement, the law defines employees as individuals with a formal employment contract with the firm.[7] This includes both formal full-time and part-time workers (i.e., salaried workers), but excludes other workers that do not have a formal employment contract, such as fee-based and commission-based workers, consultants, contractors, trainees, partners, family members, and workers hired by a third party (i.e. non-salaried workers). That is, the mandatory profit sharing requirement applies only to salaried workers, and non-salaried workers are not counted towards the 20 employee threshold.

In addition to the profit sharing requirement, there are two other regulations that apply to firms with more than 20 employees and can increase firm labor costs.[8] First, the creation of a

[4]Certain legal types of firm are also exempt, such as cooperatives, non-profits, and religious associations.

[5]Article 2 of Decreto Legislativo 892 (11.11.96).

[6]There are two mandated social security contributions in Peru. Employers must pay health contributions of 9 percent of wages to the health system (public/private) chosen by the employee. Employees must pay pension contributions of 12.4 percent to 13 percent, which employers withhold on their behalf.

[7]Article 5 of Decreto Legislativo 892 (11.11.96) and Article 2 of Decreto Supremo Nr. 009-98-TR.

[8]The regulation kink at 50 workers in France studied by Garicano et al. (2016) and Gourio and Roys (2014) also includes a mandatory profit sharing requirement. However, other important regulations change at that threshold as well, such as that management must create three worker councils, and submit restructuring plans to the councils if the company decides to fire workers for economic reasons.

firm-specific union requires at least 20 employees.[9] However, since joining a union is entirely voluntary, and there are benefits to belonging to larger unions, not all firms with more than 20 employees will have a firm-specific union. Second, the law determines that firms with more than 20 employees should set up a health and safety committee with a least four members (half appointed by workers, and half appointed by management).[10] This committee meets once a month, and meetings take place during working hours.

3 Data and stylized Facts

We use two datasets in our analysis. The first is the Encuesta Nacional Económica 2015 (ENE), a nationally representative survey of formal firms conducted by the Peruvian National Statistics Office (INEI).[11] This survey refers to the year 2014, and includes 19,204 firms with annual sales larger than 76,000 Peruvian soles (roughly $ US 20,000), or about six percent of all firms above that sales threshold. While micro firms with sales less than 76,000 soles represent more than four-fifths of all formal firms, their exclusion from the survey is not problematic for our analysis since very few micro firms have more than twenty workers, and hence will not be subject to the mandatory profit sharing or other size-dependent labor regulations.[12]

The second dataset comes from the Peruvian tax authority (SUNAT), and is compiled from information submitted by firms in their tax and social security returns. Similar to the firm survey, this dataset also refers to the year 2014 and excludes the public sector and micro firms with sales less than 76,000 soles. The dataset includes 313,810 firms with a unique taxpayer number, and thus corresponds to the near universe of firms in Peru that are affected by the profit sharing regulation.

The two datasets are complementary. While the ENE survey has a smaller sample size, it nonetheless contains more detailed information on firm sales, production, costs and employment. In particular, the ENE provides information on the number of workers by firm disaggregated between salaried and non-salaried workers. The administrative data from SUNAT includes many more firms, but the information available is limited. In particular, it does not

[9] Article 14 of Decreto Supremo Nr. 010-2003-TR.

[10] Article 29 of Law 29783

[11] More information is available on INEI's website: `http://webinei.inei.gob.pe/anda_inei/index.php/catalog/340`

[12] According to SUNAT, there were 1,647,529 registered firms in 2014, of which 1,309,399 reported having annual sales of less than 76,000 soles.

provide information on non-salaried employment by firm.

3.1 Firm size distribution

We start by examining the distribution of employment across firms in Figure (1). The figures in the top panel use data from the ENE firm survey, while the figures in the bottom panel use the administrative dataset from SUNAT, the Peruvian tax authority.

The figures on the left-hand side show the number of firms by employment bracket for firms between 10 and 30 employees. Both data sources show a discontinuity around the cut-off of 20 employees, with more firms than expected having just under 20 employees, and fewer firms than expected having 21 employees or more. The discontinuity is sharper in the ENE firm survey dataset where we have information disaggregated between salaried and non-salaried employment. By contrast, in the SUNAT dataset our employment variable includes both salaried workers and some types of non-salaried workers such as trainees, which likely explains why there is a bulge in the number of firms with fewer than 20 employees, rather than a more clear discontinuity right at the 20 employee threshold.

The figures on the right-hand side offer a wider view of the firm size distribution, by plotting the number of firms against employment brackets using a log-log scale. Overall, the firm size distribution in Peru follows a power law, similar to that in other countries (e.g. Hsieh and Klenow (2014) looking at the U.S., India, and China).[13] However, the slope of the firm size distribution appears to flatten as it approaches the 20 employee cut-off, consistent with firms bunching before that cut-off. After 20 employees, there is an intercept shift in the firm size distribution, which then resumes its approximate power law distribution for larger firms.

Overall, both datasets suggest that the profit sharing requirement may have an important impact on hiring, specifically by discouraging firms closer to the threshold from hiring more than 20 salaried employees.

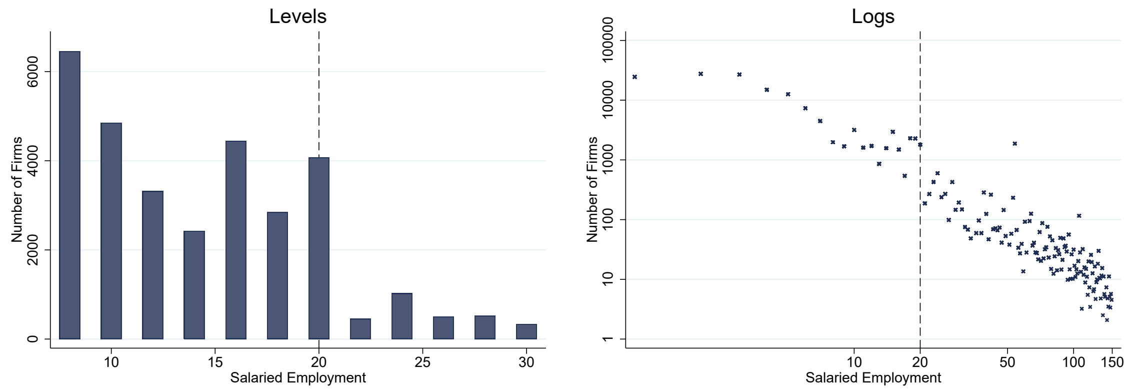
3.2 Salaried and non-salaried employment

As discussed above, the requirement to share profits is imposed by law on all firms with more than 20 salaried employees. This opens the door for firms to use informal, non-salaried workers

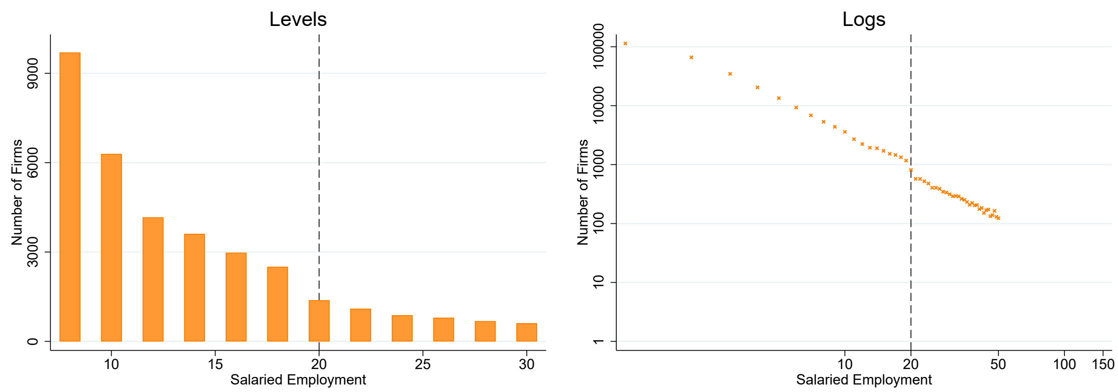
[13]Micro firms are excluded from the ENE, which explains the difference in slope of the firm size distribution in the 0-5 employee range compared to the SUNAT data.

Figure 1: Firm size distribution

(a) ENE



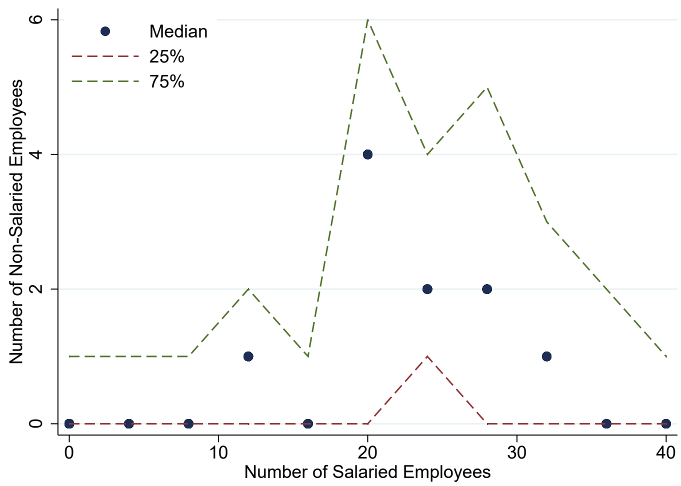
(b) SUNAT



Note: Firm employment is defined as the annual average of monthly employment, rounded to the nearest integer. Employment is presented in bins of two employees. Data refer to year 2014. SUNAT data for firms with more than 50 employees is missing due to data restrictions necessary to preserve data confidentiality.

to expand production while getting around the profit sharing requirement. Nonetheless, the productivity of salaried workers differs from that of non-salaried workers (Busso et al., 2012). This difference is explained by the nature of their position, not because of any underlying characteristic of the workers themselves. Formal workers' contracts require them to follow a specific schedule and directives by the firm's manager. This allows the firm to organize the

Figure 2: Employment composition



Note: Firm employment is defined as the annual average of monthly employment, and is presented in bins of four employees. The data are from the ENE survey, and refer to the year 2014.

production process, monitor quality, control inventories, and to coordinate tasks performed by different workers. This same level of subordination is not embedded in the contracts of non-salaried workers (e.g., fee-based or service provision contract, consultants, family members), which makes monitoring and control more difficult. Also, anecdotally, non-salaried workers have a higher turnover. This implies that non-salaried workers can be less productive, and that they are hired to perform a more limited range of tasks.

Figure (2) examines to what degree firms of different size employ non-salaried labor. On the horizontal axis, firm size is measured as salaried employment, since that is the relevant concept of employment for the profit sharing threshold. The figure plots the median and interquartile range of non-salaried employment by firm size, which is defined in consecutive four salaried worker bins to reduce sample noise. We start by observing that firms with fewer than 20 salaried workers, which are those not subject to the profit sharing requirement, tend not to hire non-salaried workers. However, the number of non-salaried workers increases discontinuously once firms reach the 20 salaried worker cutoff, with the median number of non-salaried workers increasing from zero to about one-fifth of firm employment. This is consistent with firms bunching at the 20 salaried worker policy threshold, and choosing to expand their workforce using informal labor. The number of non-salaried workers then decreases as firms hire more salaried employees, since the incentive to hire informal labor decreases once firms have met the profit sharing cut-off. The fact that some non-salaried workers are still hired even by those firms could be explained by the type of worker (e.g., outside consultants, lawyers, accountants) and potentially by frictions in adjusting employment composition in growing firms.

3.3 Sales, profits and wages

Next, we examine differences in firm performance around the 20 employee policy cut-off by focusing on median firm sales, wages and profits by employment bracket. Annual sales are taken from value added tax returns, or, in the case of smaller firms, from the simplified tax regime of sales taxation. Profits are taken from corporate tax returns, and are measured as profits before tax, dividends and profit sharing distributions. All variables are taken from the SUNAT dataset and expressed in thousands of Peruvian soles.

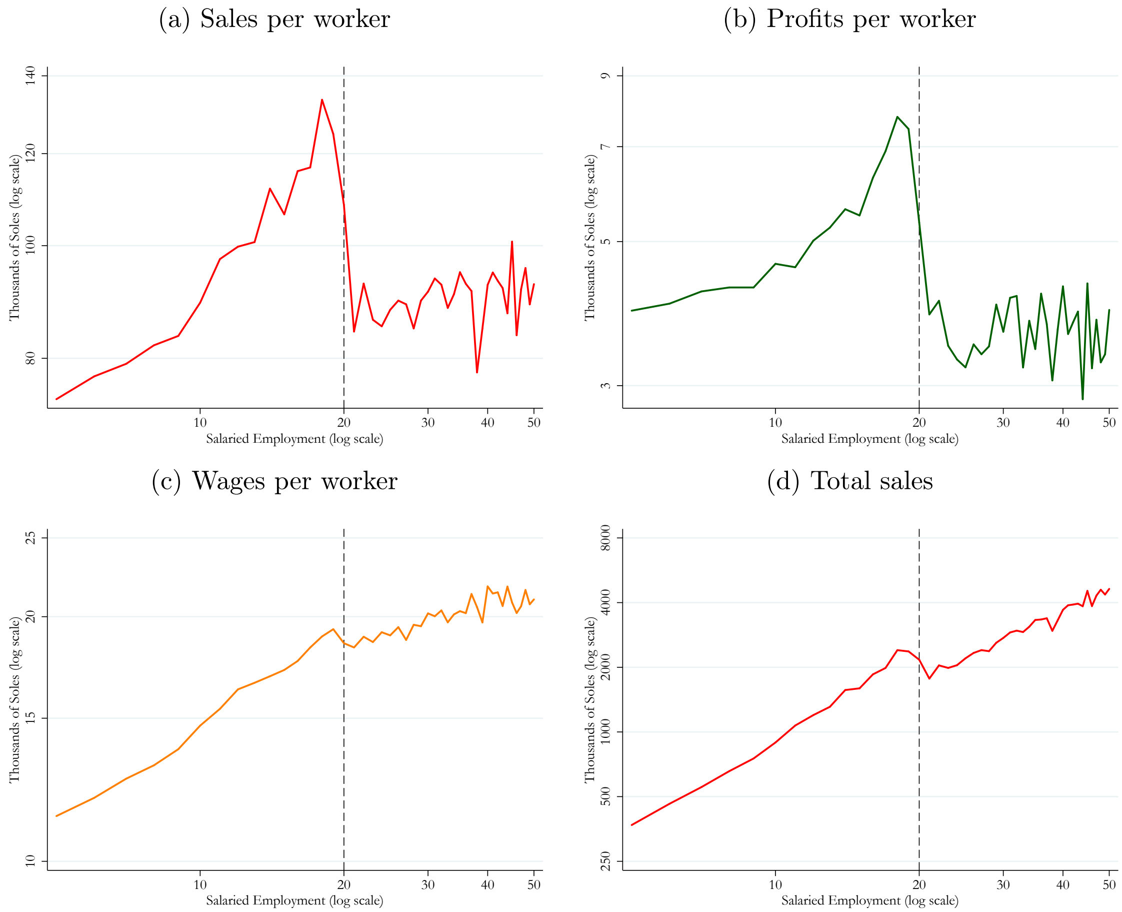
Looking at the top two panels in Figure (3), we observe clear bunching in both sales and profits per worker around the 20 employee cut-off. In particular, firms with under 20 employees appear to have much larger sales and profits per worker than would be expected given the size of their workforce.

There are several potential explanations for this sharp discontinuity. First, the measure of salaried employment we take from SUNAT excludes most informal workers, which suggests that it may underestimate total firm employment. This would then increase measured sales and profits per salary employee, making firms with under 20 salaried employees appear more productive than they truly are. However, notice that these missing workers are unlikely to be non-salaried workers, since as previously shown in Figure (2) we only observe a discontinuity in the share of non-salaried workers for firms with 20 salaried employees.

A competing explanation is that many Peruvian firms split into sub-units of less than 20 salaried employees to circumvent the profit sharing rule and other size dependent regulations, each unit with separate legal standing but under the same management. While there is ample anecdotal evidence for this phenomenon, we have no direct evidence of firm splitting in our data, since we lack information identifying firm owners.

Another explanation for this difference in profitability by firm size is that firms with less than 20 salaried employees may increase average hours per worker, or hire more qualified workers.

Figure 3: Firm performance



(a) Sales per worker

(b) Profits per worker

(c) Wages per worker

(d) Total sales

Note: Firm employment is defined as the annual average of monthly employment, rounded to the nearest integer. The data are from SUNAT, and refer to the year 2014. The figures plot median firm sales, profits, and wage bill per worker and median total sales by employment bracket. All values are expressed in thousands of Peruvian soles.

However, in the bottom left panel of Figure (3) we show that there does not appear to be a similar discontinuity in wages per worker, which would have been expected if hours worked or worker quality had changed. We do find a flatter wage profile for firms with more than 20 salaried employees, which is consistent with part of their total compensation now taking place in the form of shared profits.

Finally, the observed discontinuity may reflect higher TFP or capital intensity in firms with between 10 and 20 salaried employees. For example, Garicano et al. (2016) show some evidence of bunching in the TFP distribution when analyzing a similar policy threshold in France. Our datasets do not have an adequate measure of firm capital, so we cannot rule out that this possibility. However, it seems unlikely that higher manager productivity or capital intensity in these relatively small firms can entirely explain a difference of nearly 40 percent in sales per worker.

4 Model

We extend the model by Garicano et al. (2016), which builds on Lucas (1978), to include the firm's choice between hiring salaried and non-salaried labor. In our economy, there is a continuum of agents of measure one. Each agent is born with a managerial productivity $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, where α is distributed according to the density $\psi(\alpha)$. Knowing their productivity, agents choose whether to become a worker or to manage a firm. Workers are all identical, and are paid a wage w (or an equivalent compensation, as discussed below), regardless of their managerial productivity. Managers, who are also the firm owners, are paid the profit made by the firm that they manage. There are no costs related to starting a firm, so an agent chooses her occupation (worker or manager) by comparing the wage, w, with the profit she would make in case she managed a firm – which depends on her managerial productivity.

4.1 Firm's problem

In the absence of any size-dependent regulation, a manager with managerial productivity α makes a profit of

$$\pi(\alpha) = \max_{\ell,n} \alpha g(\ell + \phi n) - w(1+\tau_w)\ell - wn,$$

where the price of the output is normalized to one, ℓ is the number of salaried workers the firm hires and n is the number of non-salaried workers hired by the manager/firm. In our empirical analysis, we will use the "span of control" production function, $g(x) = x^\theta$ for some $\theta \in (0,1)$. However, for most of our theoretical results it suffices that that $g(x)$ is strictly increasing and strictly concave for $x > 0$.

We assume that the output generated by a non-salaried worker is only a fraction, ϕ, of the output generated by a salaried worker. Motivated by the findings in Figure (2), we assume that $\phi(1+\tau_w) < 1$. This ensures that firms will not hire any non-salaried workers unless they are constrained by a size-dependent policy[14].

On the other hand, workers are indifferent between becoming salaried or non-salaried employees, since both are paid the same wage – despite the fact that non-salaried employees produce a smaller output for the same amount of labor. Anecdotally, non-salaried employees tend to be hired for shorter spans of time, often do not have a well-defined set of tasks to accomplish, and enjoy less monitoring than salaried workers do. We interpret these facts as suggesting that non-salaried employees are less productive because of the nature of their position, and not because of any underlying characteristic of the worker herself.

Albeit not usual, there is empirical evidence that supports this assumption[15]. As already mentioned, we do not see any discontinuity in the total wage bill in the bottom left panel of Figure 3. Similarly, Figure A.1 in the appendix shows no significant difference in the average wage received by employees working in firms with 16-25 salaried employees, using data from the ENE. If the ability of salaried and non-salaried workers was fundamentally different, wages received by each type of worker would reflect this difference. Since there is a discontinuous increase in the number of non-salaried workers hired by firms with 20 salaried employees, we would also expect a discontinuous change in the average wage received by workers in firms just above and below the 20 salaried employee threshold. On the other hand, if salaried and non-salaried workers are able to perform the same tasks given the same incentives, competition would equate their wages.

The firm's optimal choice of labor is defined by the first-order condition

$$\alpha g'(\ell) = w(1 + \tau_w),$$

which defines $\ell(\alpha) = [g']^{-1}\left(\frac{w(1+\tau_w)}{\alpha}\right)$ and $n(\alpha) = 0$ as the optimal labor demands.

[14]The assumptions described in this section are made so that our model matches the stylized facts in the Peruvian economy, notably from figure 2. There are other ways in which we could do so, for example by assuming a different cost structure when hiring salaried or non-salaried workers as Ulyssea (2018) does. However, in most cases the economic content of the model would be the same, so none of the specific assumptions we make are driving the results.

[15]From a modeling perspective, the key implication of this assumption is that firms cannot fully adjust the wage of a non-salaried employee to compensate for its smaller productivity. Another wage structure that still produces this same implication would not change the interpretation of our results in any significant way.

4.2 Introducing size-dependent policies

As discussed in section (2), there are two main effects of the size-dependent policies in place in Peru. The first one is a decrease in the profits of firms, driven by the profit-sharing rule. We model the rule as a tax on profits, τ_π, that only applies to firms who hire more than some threshold, N $(= 20)$, salaried workers. The second effect are higher labor costs, driven by the mandatory institution of a health and safety committee and the potential unionization of workers in larger firms. We collective model these regulations as a tax τ_p on salaried labor.

Because workers in firms with more than N salaried employees are getting a share of the firm's profits, a reasonable response by firms is to reduce the wages paid to those workers. Let $c(\tau_\pi, \tau_p)$ be the compensation received by salaried employees who are entitled to a share of the firm's profits. Under the size dependent policies, a manager with managerial productivity α will have profits described by

$$\pi(\alpha) = \max_{\ell,n} \left\{ \begin{array}{ll} \alpha g(\ell + \phi n) - w(1+\tau_w)\ell - wn, & \text{if } \ell \leq N \\ (1-\tau_\pi)[\alpha g(\ell + \phi n) - (1+\tau_w+\tau_p)c(\tau_\pi, \tau_p)\ell - wn], & \text{if } \ell > N \end{array} \right\}$$

We show in appendix B that, under the assumption that workers' indirect utility function is multiplicatively separable (this is true, for example, for CRRA utility functions), in equilibrium the compensation $c(\tau_\pi, \tau_p)$ is proportional to the wage w. Because of this, we can rewrite $(1 + \tau_w + \tau_p)c(\tau_\pi, \tau_p) = Tw$, where T is a re-scaled labor tax. Once again motivated by the findings in Figure (2), we assume that $T\phi < 1$, so that large firms who are actively hiring salaried workers will not hire non-salaried workers.

We end this section pointing out some potential caveats in our model's environment. While firms can reduce the wages of salaried workers who receive a share of the profits, removing any gain that the profit-sharing rule might impart to workers, the ability to unionize or having a health and safety committee on the work place could be beneficial to workers. We do not include those benefits in our model for two reasons. First, they are subjective and potentially very different across workers. This makes them hard to measure and to quantify in a model. Second, for any of those benefits to materialize, we would need to include some other friction into the model. One example would be search frictions that lead workers and firms to bargain on wages (so that unionizing could increase the bargaining power of workers), which is outside the scope of the present paper. Absent those frictions, there is nothing that stops firms from once more adjusting wages to compensate for the added benefits of the regulation. Precisely because workers in large and small firms are identical, in equilibrium their payoffs must be identical as well. This simple reasoning removes any potential gains to workers that might come from a size-dependent policy.

Finally, one could argue that the profit-sharing rule could actually be beneficial to firms, since it helps to mitigate moral hazard problems. However, if it were the case that firms were benefitting from this rule, then firms not subject to the regulation could mimic its effects by engaging in profit sharing as well. Instead, what we observe is that smaller firms do not engage in profit-sharing. Moreover, firms that are close to the profit-sharing threshold go to great lengths to avoid having to share their profits with employees, which suggests that there is little benefit from doing so.

4.3 Characterizing the labor demands

Agents' choices in this model depends on their managerial productivity. To characterize the labor demands for different firms, it is useful to define a series of thresholds, graphically represented in Figure 4.

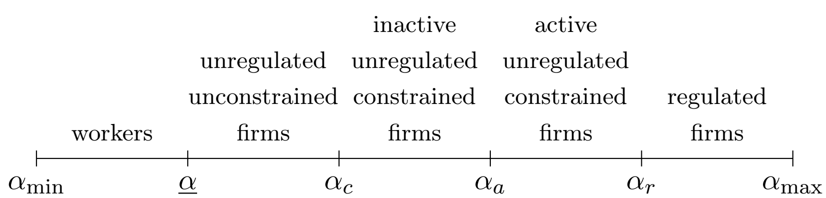


Figure 4: Graphical representations of the many thresholds in the model.

The first relevant threshold for firms is α_c, the managerial productivity of a manager who hires exactly N salaried workers regardless of the size-dependent policies. Managers with $\alpha \leq \alpha_c$ are unconstrained by the size-dependent regulations, and their labor demand is characterized above. It follows that

$$\alpha_c = \frac{w(1+\tau_w)}{g'(N)}.$$

Because of the discontinuous drop in profits that happens when firms hire more than N salaried employees, managers whose ability is $\alpha > \alpha_c$ will have incentives to hire exactly N salaried employees, and to supplement their labor force with non-salaried employees. A firm in that situation would solve

$$\max_n \alpha g(N + \phi n) - w(1 + \tau_w)N - wn.$$

The FOC is

$$\alpha g'(N + \phi n(\alpha))\phi = w$$

and therefore

$$n(\alpha) = \frac{1}{\phi}[g']^{-1}\left(\frac{w}{\alpha\phi}\right) - \frac{N}{\phi}.$$

However, note that $n(\alpha) \geq 0$ is true if, and only if,

$$[g']^{-1}\left(\frac{w}{\alpha\phi}\right) \geq N$$

$$\alpha \geq \frac{w}{\phi g'(N)} = \frac{\alpha_c}{\phi(1 + \tau_w)} = \alpha_a$$

where we have used that g' is strictly decreasing to switch the sign of the inequality. It follows that for $\alpha \in [\alpha_c, \alpha_a)$, managers will optimally choose to be "inactive" in the sense that they are constrained by the size-dependent policy but decide not to hire non-salaried workers. On the other hand, managers whose $\alpha \geq \alpha_a$ will be "active". They are constrained by the size-dependent policy to hire only N salaried workers, but will hire non-salaried workers to supplement their work force.

Non-salaried workers are less productive than salaried workers ($\phi < 1$), which means that it is not optimal for all managers whose managerial ability is greater than α_c to hire only N salaried workers. As α increases, managers demand more labor, and there is a point where coping with the size-dependent policies costs and hiring more salaried workers will be more profitable than to keep hiring non-salaried workers. This point is determined by α_r, which is the managerial ability of the manager who is indifferent between coping with the regulation costs and hiring the optimal amount of salaried workers or hiring only N salaried workers to get around the regulation, while supplementing their labor demand with non-salaried workers. Hence, α_r is implicitly defined by

$$(1 - \tau_\pi)[\alpha_r g(\ell(\alpha_r)) - Tw\ell(\alpha_r)] = \alpha_r g(N + \phi n(\alpha_r)) - w(1 + \tau_w)N - wn(\alpha_r)$$

Finally, since $T\phi < 1$, all managers with $\alpha \geq \alpha_r$ will never hire non-salaried workers, and their labor demand is defined by

$$\ell(\alpha) = [g']^{-1}\left(\frac{wT}{\alpha}\right) \quad \text{and} \quad n(\alpha) = 0$$

4.4 Equilibrium

To fully characterize the choices made by agents in this economy, we are only missing one threshold that defines which agents become managers versus workers. This threshold, $\underline{\alpha}$, is defined by

$$\underline{\alpha} g(\ell(\underline{\alpha})) - w(1+\tau_w)\ell(\underline{\alpha}) = w \tag{1}$$

that is, the manager whose managerial productivity is $\underline{\alpha}$ is indifferent between managing a firm and working for someone else. Since w is fixed, it is straightforward to see that any agent whose $\alpha < \underline{\alpha}$ is strictly better off being a worker, while any agent whose $\alpha > \underline{\alpha}$ is strictly better off being a manager. This cutoff also defines the labor demand of the smallest firm in the economy, $\ell_{\min} = \ell(\underline{\alpha}) = [g']^{-1}\left(\dfrac{w(1+\tau_w)}{\alpha}\right)$.

Gathering all of our derivations so far, the labor demand as a function of α can be written as

$$\ell(\alpha) = \begin{cases} 0, & \text{if } \alpha \in [\alpha_{\min}, \underline{\alpha}) \\ [g']^{-1}\left(\frac{w(1+\tau_w)}{\alpha}\right), & \text{if } \alpha \in [\underline{\alpha}, \alpha_c] \\ N, & \text{if } \alpha \in (\alpha_c, \alpha_r) \\ [g']^{-1}\left(\frac{wT}{\alpha}\right) & \text{if } \alpha \in [\alpha_r, \alpha_{\max}] \end{cases} \tag{2}$$

and

$$n(\alpha) = \begin{cases} 0, & \text{if } \alpha \in [\alpha_{\min}, \alpha_a] \\ \frac{1}{\phi}[g']^{-1}\left(\frac{w}{\alpha\phi}\right) - \frac{N}{\phi}, & \text{if } \alpha \in (\alpha_a, \alpha_r) \\ 0 & \text{if } \alpha \in [\alpha_r, \alpha_{\max}] \end{cases} \tag{3}$$

The market clearing condition in the labor market is

$$\int_{\alpha_{\min}}^{\underline{\alpha}} \psi(\alpha) d\alpha = \int_{\underline{\alpha}}^{\alpha_{\max}} (\ell(\alpha) + n(\alpha))\psi(\alpha) d\alpha, \tag{4}$$

with $\ell(\alpha)$ and $n(\alpha)$ are defined in 2 and 3. Because there is only one other market (for the consumption good), this condition is enough to characterize an equilibrium.

Definition: Given a distribution of managerial productivity $\psi(\alpha)$ over $[\alpha_{\min}, \alpha_{\max}]$, a production function $y = \alpha g(\ell + \phi n)$, taxes τ_w, τ_p and τ_π, a competitive equilibrium in this economy consists of: (i) a wage w^*; (ii) a labor allocation $\ell^*(\alpha)$ and $n^*(\alpha)$ that assign ℓ^* salaried and n^* non-salaried workers to a firm whose manager productivity is α; and (iii) a set of cutoffs $\{\underline{\alpha}, \alpha_c, \alpha_a, \alpha_r\}$ where $[\alpha_{\min}, \underline{\alpha})$ is the set of workers, $[\underline{\alpha}, \alpha_c]$ is the set of unregulated and unconstrained firms; $(\alpha_c, \alpha_a]$ is the set of inactive, unregulated and constrained firms; (α_a, α_r) is the set of active, unregulated and constrained firms; and $[\alpha_r, \alpha_{\max}]$ is the set of regulated firms. These objects are defined such that no agent wishes to change occupation (equation 1), managers hire the number of employees that maximize their profits given w^* and taxes (equations 2 and 3), and the labor market clears (equation 4).

5 Extensions

5.1 Firm splitting

We also allow for the possibility that firms split into two or more subfirms to avoid being regulated.[16] To be able to model firm splitting, we need to make some assumptions on the way that firms split. The three assumptions listed below are sufficient for us to derive (and estimate) the firm size distribution when firms are allowed to split. This is done in Appendix E.

Assumption 1: Each firm has a probability $\delta \in (0, 1)$ of splitting. Furthermore, δ is independent of the firm's characteristics.

This first assumption is driven by the fact not all firms split. Otherwise, all firms whose managers have productivity $\alpha > \alpha_c$ would simply divide their firms into smaller units, and we would not observe any large firms in the data. Assuming that firms are not able to split with

[16] Onji (2009) finds similar evidence of firm splitting for Japan, in the context of minimum VAT thresholds.

probability $1-\delta$ is the simplest way of matching this fact in the data. The assumption that δ is independently distributed across firms is not very restrictive, as it simply requires that the decision to split be unrelated to firm size. This decision could still be driven by a number of factors not included in the model, including the firm's sector, local regulations, and personal preferences of the manager.

Assumption 2: If a firm splits, it will divide itself in the smallest possible number of sub-firms (as long as all sub-firms have less than N salaried employees).

This assumption is justified by the fact that splitting a firm involves costs (forgone economies of scale, potentially a separate site of operation, a new set of accounting responsibilities, etc.) that scale up with the number of units that a firm splits into. We can think of the cost of splitting as being a small but positive fixed cost incurred for each new unit of a firm. This cost is unlikely to affect anything else in the model, but would push firms into splitting into the smallest number they can. For example, a firm with 64 employees could split into 8 sub-firms with 8 employees each, or 4 sub-firms with 16 employees each. Assumption 2 implies that this firm would split into 4 sub-firms.

Assumption 3: Firms split into units of equal size.

Finally, to be able to model firm splitting, we need to know how firms typically split. In this respect, assuming that they split into units of equal size seems like a natural way to proceed. We should also make it clear that a manager who splits his firm will still have the same production function and the same span of control over his employees – e.g. a manager who manages one firm with 30 employees produces the same as an equally productive manager who manages two firms with 15 employees each. Finally, we continue to assume that firms can hire a non-integer number of employees, but we will restrict the number of units that a firm splits into to be an integer number.

5.2 Partially rigid wages

Like Garicano et al. (2016), we consider both the case where wages are flexible, and the case where wages do not fully adjust after the regulations are imposed. Partially rigid wages are an important feature since rigidities in the Peruvian labor market, including the minimum wage, likely create frictions that prevent a full downward adjustment of wages for formal workers. Thus, if wages do not fully adjust once the regulations are put into place, this may amplify their impact on overall employment and income.

We model wage inflexibility following Garicano et al. (2016). They assume that after a regulatory change the observed wage w is a weighted average of the original wage w_0 (before the regulatory change), and the "target" wage w^* (i.e. the wage level that would equilibrate markets with no unemployment):

$$w = \rho w^* + (1 - \rho) w_0. \tag{5}$$

In Appendix F, we discuss how to calibrate ρ to match differences in unemployment between Peru and similar countries with less stringent size-dependent policies.

If $\rho < 1$, the economy will feature an unemployment rate $u > 0$. The model with partially rigid wages works just as the one described above, with two key differences. The labor market clearing equation becomes

$$(1 - u) \int_{\alpha_{\min}}^{\underline{\alpha}} \psi(\alpha) d\alpha = \int_{\underline{\alpha}}^{\alpha_{\max}} (\ell(\alpha) + n(\alpha)) \psi(\alpha) d\alpha, , \tag{4'}$$

where the supply of labor now accounts for unemployed workers; and the sorting condition 1 becomes

$$\underline{\alpha} g(\ell(\underline{\alpha})) - w(1 + \tau_w)\ell(\underline{\alpha}) = w(1 - u) \tag{1'}$$

where $w(1 - u)$ is the expected wage of a worker. The worker makes a wage w is employed and 0 if unemployed.

6 Model estimation

6.1 Empirical implications of the model

The key empirical implication of our model that we will explore in the data is the shape of the firm size distribution. Figure 5a shows the labor demands as a function of managerial ability, as described in equations 2 and 3. Using the fact that $\alpha \sim \psi(\alpha)$, we can construct the distribution of firm size using these labor demands.

It is a well-known empirical regularity that the firm size distribution approximately follows a power law. Given $g(x) = x^\theta$, Lucas (1978) shows that managerial productivity must also follow a power law if we want to be consistent with this empirical regularity. We therefore assume $\psi(\alpha) = C\alpha^{-\beta}$, for some $C, \beta > 0$. Using the change of variable formula, the firm size distribution is

$$\zeta(L(\alpha)) = \frac{1}{P}\psi(\alpha(L))\alpha'(L)$$

where P is the share of managers in the economy and L stands in for salaried, non-salaried or total (salaried + non-salaried) employment. $\alpha(L)$ is the inverse of the demand functions defined in equations 2 or 3. Figure 5b plots the log-density of total employment (and consequently also the density of salaried employment).

In our empirical applications, we will mostly focus on the distribution of salaried employment, since it already identifies all the model's parameters. Moreover, salaried employment is likely to be better measured than non-salaried employment, which some firms may choose not to report. The distribution of firm sizes, as measured by salaried employment, is given by (see Appendix D.1 for the derivation)

$$\zeta(\ell) = \begin{cases} \frac{C}{P}\left(\frac{\theta}{(1+\tau_w)w}\right)^{\frac{\gamma-1}{1-\theta}}(1-\theta)\ell^{-\gamma}, & \ell_{\min} \leq \ell < N \\ \frac{C}{P}\left(\frac{\theta}{w}\right)^{\frac{\gamma-1}{1-\theta}}\left(\frac{1-\theta}{\gamma-1}\right)\left[(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}N^{1-\gamma} - T^{\frac{1-\gamma}{1-\theta}}\ell_r^{1-\gamma}\right], & \ell = N \\ 0, & N < \ell < \ell_r \\ \frac{C}{P}\left(\frac{\theta}{Tw}\right)^{\frac{\gamma-1}{1-\theta}}(1-\theta)\ell^{-\gamma}, & \ell_r \leq \ell \leq \ell_{\max} \end{cases} \tag{6}$$

where $\gamma = \beta(1-\theta) + \theta$, $\ell_{\min} = \ell^*(\underline{\alpha})$, $\ell_r = \ell^*(\alpha_r)$ and $\ell_{\max} = \ell^*(\alpha_{\max})$.

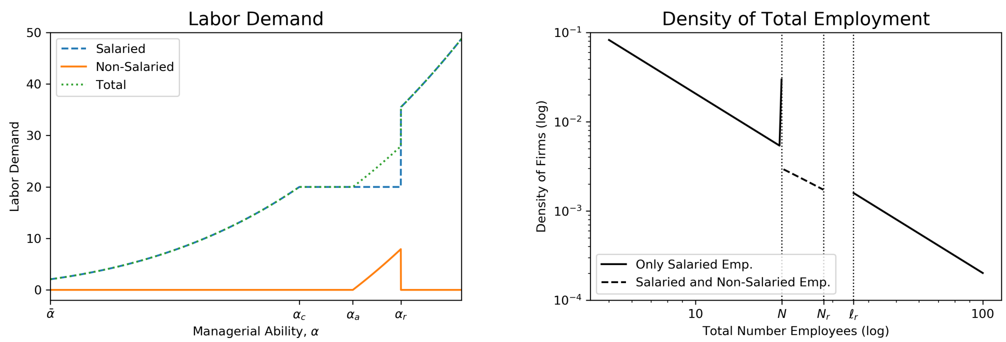
6.2 Estimation

To begin the estimation of our model, we first set $\tau_w = 0.13$, which is the cost of pension contributions as a share of the wage in Peru (Alaimo et al., 2017). Our estimation procedure follows 3 steps:

1. Estimate θ and ϕ using the production function specification;
2. Estimate γ, ℓ_r, T and $\ell_{\max}$ using the firm size distribution;

Figure 5: Model predictions for labor demands and density of total employment.

(a) Labor demands (b) Log density of total employment



Note: Panel (a) shows the labor demands as a function of α. Panel (b) plots the log density of total employment, which is defined as salaried plus non-salaried employment. Note that $N_r = n^*(\alpha_r) + N$.

3. Use model restrictions to find the remaining parameters.

Finally, we remove outliers for the estimation procedure.[17] Out of a little over 19,200 observations, 170 are flagged as outliers and removed. Due to the small number of outliers, most of our results are not sensitive to them. However, they do affect the numerical stability of parts of the estimation, in particular the maximum likelihood estimation of the firm size distribution.

6.2.1 Production function estimation

Our model assumes that labor is the only factor used in production. As a result, in our estimation of the production function, we are only interested in recovering the labor coefficients. The production function has the form

$$\ln(y_j) = \theta \ln(\ell_j + \phi n_j) + \ln(\alpha_j)$$

[17]Since we are estimating the firm size distribution, outliers are defined based on the share of firms with a given number of salaried employees. If share(X) is the share of firms with X salaried employees, outliers are defined as firms who hire X salaried employees and

$$\text{share}(X) > 10 \max\{\text{share}(X-1), \text{share}(X+1)\} \quad \text{or} \quad \text{share}(X) < \frac{1}{10} \min\{\text{share}(X-1), \text{share}(X+1)\}.$$

Table 1: Regressions of log-sales on log-inputs and controls

	(1)	(2)	(3)
Log Salaried	0.607***	0.633***	0.712***
	(0.058)	(0.095)	(0.075)
Log Non-Salaried	0.376***		
	(0.066)		
Sample	All	Salaried < 20	Salaried Only
Observations	1086	1859	1128
R-squared	0.89	0.64	0.79

Note: All specifications control for average worker education, 4-digit industry sector and firm type (e.g. personal, LLC, SA). The number of observations available to run these regressions is limited because of missing data on firm capital and capital purchases. Standard errors are robust to heteroskedacity. * 0.10, ** 0.05, *** 0.01.

for each firm j. There are two challenges in estimating θ and ϕ. The first is the non-linearity of the log-log specification. The second is that α_j is not observed, giving rise to well-known endogeneity issues or "transmission bias". To address the first problem, we estimate a number of different linear and non-linear specifications of the model, showing how both θ and ϕ can be identified from the data. To address the second problem, our regressions are analogous to the first stage of the method proposed by Olley and Pakes (1996), and include a non-parametric proxy for unobserved productivity.

To estimate θ, we run

$$\ln(\text{sales}_j) = \beta_0 + \beta_1 \ln(\text{salaried emp}_j) + \beta_2 \ln(\text{non-salaried emp}_j) + \gamma X_j + \varepsilon_j \tag{7}$$

where X_j are controls included in the regression. X_j and includes physical capital utilized by firms, indexes of worker and manager educational achievement (varying between 0 (no high school) to 1 (post-graduate)), a firm type fixed effect (e.g. L.L.C., S.A., person, etc), a 4-digit industry fixed effect, and a complete 3rd-degree polynomial on manager education and capital purchases, included as an approximation for managerial productivity. The estimation results are in Table 1.

The coefficient on salaried labor identifies θ, which is the elasticity of output with respect to to labor. Column (1) in Table 1 runs the regression specified in 7 over the entire sample of firms

Table 2: Regressions of log-sales on level of inputs and controls

# Salaried	0.074***
	(0.019)
# Non-Salaried	0.062***
	(0.005)
Implied ϕ	0.831***
	(0.107)
Observations	2013
R-squared	0.66

Note: All specifications control for average worker education, 4-digit industry sector and firm type (e.g. personal, LLC, SA). Standard errors are robust to heteroskedacity. * 0.10, ** 0.05, *** 0.01.

and reports the coefficients on labor. Column (2) runs the same specification (omitting the non-salaried labor term) for the sub-sample of firms who hire less than 20 salaried employees (and which the model predicts should not hire any non-salaried employees). Finally, column (3) presents the same regression for the sub-sample of firms who do not hire non-salaried employees. This is our preferred specification, since we can directly map the elasticity of salaried employees to θ in our model by restricting the sample to firms who do not hire non-salaried labor. However, note that the coefficients on salaried labor are economically similar and not statistically different across specifications.

Recent papers by Ackerberg et al. (2015) and Gandhi et al. (2017) point out issues with the proxy variable methodology. In particular, they show that it may not identify the production function. With that in mind, we show in Appendix C.1 how θ can be mapped into the labor cost share of firms, which is observable in our data. Calibrating θ to match the average (or median) labor cost share would give us $\hat{\theta} \approx 0.7$, which is very much in line with the values we found above.

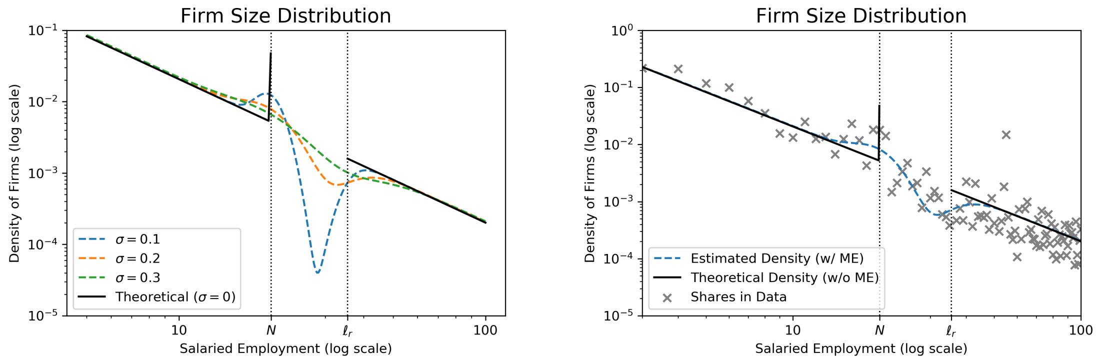
To estimate ϕ, we note that

$$y = (\ell + \phi n)^{\theta} \Longrightarrow \frac{\partial \log(y)}{\partial \ell} = \frac{\theta}{\ell + \phi n} \quad \text{and} \quad \frac{\partial \log(y)}{\partial n} = \frac{\theta \phi}{\ell + \phi n}$$

We can estimate the semi-elasticities above (more precisely, their average across our sample) by running a slightly modified version of 7, where we use the labor levels instead of logs. ϕ

Figure 6: Firm size distribution

(a) Measurement error (b) Estimated density



Note: Panel (a) shows the effects of adding measurement error to the firm size distribution. Panel (b) plots the estimated density.

is then identified by the ratio of the coefficient on non-salaried labor and the coefficient on salaried labor.

Note that we only used the sub-sample of firms who have less than 20 salaried employees to estimate those coefficients. The reason for that is that a non-trivial share of firms are constrained by the regulations in place, and therefore their employment decision are distorted. The estimation results are shown in Table 2. The standard errors for the implied ϕ are calculated using the delta-method. To check the robustness of our results, in Appendix C.2 we run two different regression specifications that identify ϕ.

Gathering the results shown in Tables 1, 2 and Appendix C.2, we use $\theta = 0.7$ and $\phi = 0.8$ when estimating the firm size distribution and performing policy counterfactuals.

6.2.2 Estimating the firm size distribution

Our estimation of the parameters in the firm size distribution follows the procedure proposed by Garicano et al. (2016). Our starting point is $\zeta(\ell)$, defined in equation 6. The theoretical distribution $\zeta(\ell)$ has two distinctive features: a mass point at N and no firms hiring between N and ℓ_r salaried workers (see Figure 5b), both not observed in the data. Garicano et al. (2016) propose that the cause of this disparity is that we only observe labor with measurement error. In particular, we only observe

$$\ell(\alpha, \varepsilon) = \ell^*(\alpha)e^{\varepsilon}$$

where $\varepsilon \sim \mathcal{N}(0, \sigma)$ is an i.i.d. shock. The advantage of adding measurement error in this way is that we can solve for the distribution of $\ell(\alpha, \varepsilon)$. First note that

$$\mathbb{P}(\ell(\alpha, \varepsilon) < x|\varepsilon) = \mathbb{P}(\ell^*(\alpha) < e^{-\varepsilon}x|\varepsilon) = \int_{\ell_{\min}}^{e^{-\varepsilon}x} \zeta(y)dy.$$

Since ε is normally distributed, we can integrate it out to find

$$\mathbb{P}(\ell(\alpha, \varepsilon) < x) = \int_{\mathbb{R}} \mathbb{P}(\ell(\alpha, \varepsilon) < x|\varepsilon)\varphi\left(\frac{\varepsilon}{\sigma}\right)\frac{1}{\sigma}d\varepsilon,$$

where φ is the normal density. Finally, the density of $\ell(\alpha, \varepsilon)$ – i.e. observed labor – is

$$\omega(x) = \frac{\partial\mathbb{P}(\ell(\alpha, \varepsilon) < x)}{\partial x}.$$

The details of the derivation are in appendix D.2.

Adding measurement error to the theoretical distribution has two effects. First, it transforms the mass point in N into a "bulge". Second, it smooths out the discontinuous drop to zero in the number of firms who hire between N and ℓ_r salaried employees. Figure 6a illustrates this point for various levels of σ.

We estimate the parameters in the firm size distribution using standard maximum likelihood. However, we do include one constraint into the estimation, which is that firms who hire ℓ_r salaried employees are not worse off than they would be by hiring N salaried workers and more non-salaried workers instead. In other words, we require that the indifference condition defined by α_r is satisfied by our estimated parameters.

Because there are 3 different profit sharing levels depending on the firm's sector, we require the weakest possible version of this constraint to be satisfied:

$$(1 - \tau_\pi^{\min})[\alpha_r g(\ell(\alpha_r)) - Tw\ell(\alpha_r)] \geq \alpha_r g(N + \phi n(\alpha_r)) - w(1 + \tau_w)N - wn(\alpha_r)$$

where $\tau_\pi^{\min} = 0.05$ is the smallest possible profit sharing level. The estimated parameters are

Table 3: Parameters from maximum likelihood estimation

Parameter	Estimated Value	95th CI
γ	1.99	[1.98, 2.00]
σ	0.182	[0.180, 0.184]
ℓ_r	35.5	[35.1, 35.9]
T	1.143	[1.142, 1.144]
$\ell_{\max}$	33451	[29255, 37647]

shown in Table 3. Figure 6b displays the fit of our estimated firm size density to the data.

To compute the counterfactuals that follow, we also need to compute the value of a few remaining parameters in the model: the thresholds $\alpha_{\max}, \alpha_a, \alpha_c, \underline{\alpha}, \alpha_{\min}$, and w, C, P, τ_π. We normalize $\alpha_{\max} = 1$ and use model restrictions to find the value of all other parameters. This procedure is described in detail on appendix D.3.

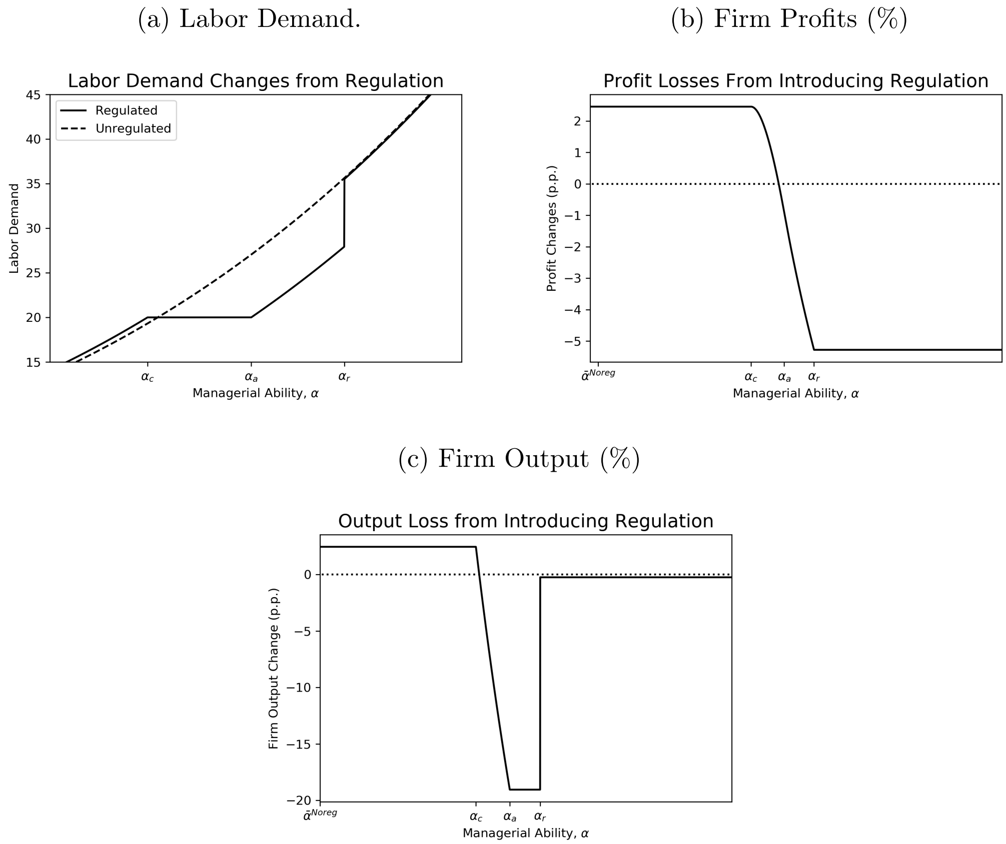
7 Counterfactuals

We consider three policy counterfactuals to assess the implications of the size-dependent regulations in Peru.[18] The first counterfactual measures the cost that these regulations impose on the Peruvian economy. We start by computing what the economy would look like if there were no regulations in place. We then introduce the regulations and measure changes in wages, aggregate production, aggregate profits, and the share of managers (or firms) in the economy. In the case where wages are not fully flexible, we also measure the effect on unemployment.

In the first counterfactual exercise, firms can hire non-salaried workers to get around the regulation. The second counterfactual removes this margin of adjustment. As pointed out throughout the paper, non-salaried employees act as a substitute for salaried employees for regulated firms, which allows those firms to reduce the regulation costs they face. Removing non-salaried labor from the economy allows us to assess how effective firms are in mitigating the negative effects of the size-dependent regulations by relying more on informal workers. Again, we focus on changes in the labor demand for each firm and at aggregate variables such as wages, aggregate production, aggregate profits and the share of managers in the economy.

[18] All of the figures that follow are computed using the model without firm splitting. The figures when firm splitting is included are similar.

Figure 7: Effects of removing size-dependent regulations



Finally, we consider the economic impact of varying the value of the threshold, rather than removing the size-dependent policies entirely. While removing the regulations may be a first-best policy option, in practice this is not always possible to implement. In contrast, changing the value of threshold may be feasible, and it is important to explore whether small changes in threshold values can deliver large economic gains.

7.1 The cost of regulation

The first counterfactual considers the economic impact when the size dependent regulations are put into place. These regulations include both the profit sharing rule and the health and union benefits that increase labor costs. As expected, the firms most affected by regulation are those medium-sized firms in the vicinity of 20 employees. This corresponds to about 3

Table 4: Aggregate impact of the size-dependent regulations (in percent)

	Aggregate Wages	Aggregate Profits	Aggregate Output	Share of Managers
Baseline Model				
Flexible wages	-1.034	-2.884	-0.096	3.504
Rigid Wages	-0.644	-3.774	-1.006	3.501
With firm splitting				
Flexible wages	-0.775	-2.928	-0.084	2.576
Rigid wages	-0.384	-3.809	-0.995	2.573

percent of firms in the Peruvian economy, who employ about 10 percent of formal workers.

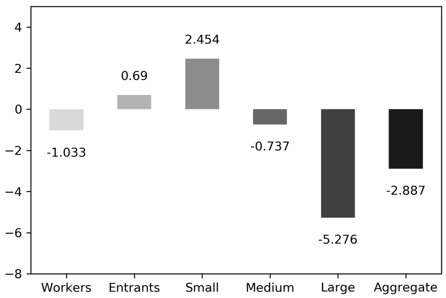
As shown in Figure 7a, these firms employ much less labor (both salaried and non-salaried) than they otherwise would as they try to minimize regulatory costs. Smaller firms (i.e. those with less than 20 employees) actually benefit slightly from the introduction of the size dependent regulation, as the decrease in overall labor demand leads to lower average wages, allowing small firms to increase employment. Figure 7b makes this point even clearer, showing that small firms' profits increase by about 2% when the regulations are put into place. On the other hand, because of the profit sharing rule, larger firms' (after tax) profits decrease by a little over 5%. Finally, Figure 7c shows the variation in output of each type of firm. Note that, because they cut back on the number of workers, firms whose manager's ability falls into the range (α_a, α_r) experience a drop of almost 20% in their output.

Table 4 presents the aggregate impact of introducing the size-dependent regulations under different versions of the model, including fully flexible and partially rigid wages, and with and without firm splitting. While salaried workers in larger firms would receive higher total compensation thanks to the profit sharing rule, in the aggregate, workers are worse off with the regulation because lower demand for labor drives wages down. This decline is larger than 1 percent in the case that wages are fully flexible and there is not firm splitting, but close to 0.4 percent when there is some downward wage rigidity and firms can split.

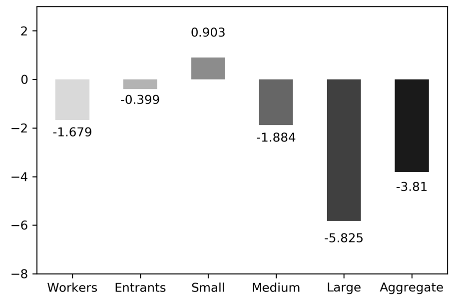
At the same time that wages decrease, small firms' profits increase, making it more attractive to become an entrepreneur. As a result, the share of managers in the economy increases, which goes along with an increase in the share of small firms and a decline in average firm

Figure 8: Winners and losers from introducing regulations
(% change in income)

(a) Baseline Model.



(b) With Firm Splitting and Rigid Wages



Note: Groups are defined based on their managerial productivity. Workers are agents who choose to be workers both before and after the regulation is put in place. Entrants are agents who were workers before the regulation but become firm owners. Small firms refer to firms owned agents who choose to be managers both with and without the regulation, but whose managerial ability is $\alpha \leq \alpha_c$ (i.e. they would *optimally* hire less than 20 workers with the regulation in place). Medium firms are firms whose manager has ability between α_c and α_r (i.e. those firms who hire only 20 salaried workers when the regulation is place, but who would like to hire more). Large firms are firms whose manager has ability $\alpha \geq \alpha_r$ (i.e. who hire more than 20 salaried workers even under the regulation regime). The change in worker's income in the right panel does not match the variation in wages in Table 4 because it takes into account the probability that a worker will be unemployed, which lowers their expected income.

productivity. Moreover, aggregate profits are lower by about 3 to 4 percent, as losses for larger firms more than offset the gains by smaller firms and the fall in wages.

The impact on aggregate production greatly depends on the degree of wage flexibility. In the case of flexible wages, output decreases only marginally, by about 0.1 percent, since just a few firms see sizeable changes in their output. However, when wages are partially rigid, the impact on output is much larger, with output declining by about 1 percent of GDP. Intuitively, the introduction of the size-dependent regulations lowers labor demand, lowering wages. In the presence of wage rigidities, wages cannot fully adjust, leading to a larger decline in employment in equilibrium. The presence of wage rigidity therefore generates unemployment, as wages remain too high for markets to clear. In our calibration (see Appendix F), we calculate that unemployment increases by 1.3 percent when the regulations are implemented and wages are not fully flexible. For comparison, the unemployment rate in Peru in 2014 was 5.5 percent, suggesting that these size-dependent regulations may account for about one fourth of total structural unemployment in the country.

The results when firms can split are similar. Wages now fall by less, since firm splitting moderates the decrease in labor demand. Aggregate profits and output would fall by roughly the same amount as when firm splitting is not allowed, since firms that split are able to avoid profit sharing but at the same face a higher cost of labor. There is a smaller increase in the share of firms in the economy, and therefore a smaller decrease in average firm productivity, since some larger firms are able to split and continue growing.

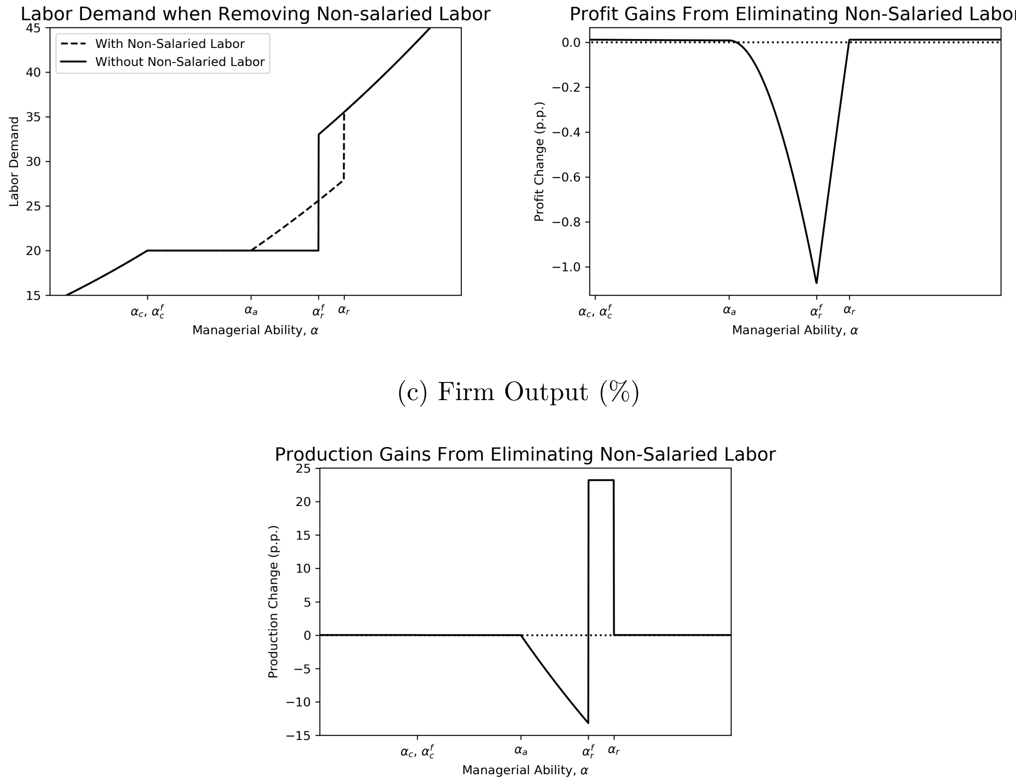
Figure 8 shows the winners and losers from introducing the size-dependent regulations in the economy. Each bar represents the change in income (wages/profit) for a specific group of the population. Because the sizes of firms change when the regulations are introduced, the population groups are defined based on their managerial productivity. The groups that benefit from the regulation are small firms and entrants (meaning those agents who start small firms once the regulations are introduced) because they face a lower overall wage, and thus are able to expand employment and increase profitability. However, these improvements come at the expense of larger firms and workers. Larger firms are negatively affected, especially through a severe reduction in profitability. Meanwhile, workers face lower wages and higher unemployment.

7.2 Removing non-salaried labor

The second counterfactual we consider is the removal of non-salaried labor from the economy. When we remove non-salaried labor, only firms in the "active, unregulated and constrained" regions in Figure 4 are directly affected. For these firms, there are two counteracting effects on their labor demands. The first effect pushes down the demand for labor, as constrained firms hire less labor because they are no longer allowed to hire non-salaried employees. This pushes these firms into only hiring N salaried employees and nothing more. The second effect pushes up the demand for labor. As shown in Figure 9a, the manager who is indifferent between complying or avoiding the regulations has productivity $\alpha_r^f < \alpha_r$, which implies that managers return to hiring salaried employees earlier, increasing the demand for labor. Intuitively, not being able to hire non-salaried workers leads to lower profits, so the more productive constrained firms now have a stronger incentive to comply with the regulations and resume hiring salaried employees. As it happens, both effects almost cancel each other out, and there is little change in the aggregate demand for labor.

Panels 9b and 9c show a similar story. Only active constrained firms are affected by the

Figure 9: Effects of removing non-salaried labor



(a) Labor Demand

(b) Firm Profits (%)

(c) Firm Output (%)

Table 5: Aggregate impact of eliminating non-salaried labor (in percent)

	Aggregate Wages	Aggregate Profits	Aggregate Output	Share of Managers
Baseline model				
Flexible wages	-0.005	0.002	0.040	0.017
Rigid Wages	-0.003	-0.002	0.035	0.017
With firm splitting				
Flexible wages	-0.003	-0.002	0.036	0.010
Rigid wages	-0.001	-0.006	0.032	0.010

Figure 10: Winners and losers from removing non-salaried labor
(% change in income)

(a) Baseline Model. (b) With Firm Splitting and Rigid Wages

Note: See note to Figure 8.

removal of non-salaried labor. Other firms could be indirectly affected by changes in wages, but as discussed below, the variation on wages is very small. It is nonetheless interesting to look at the large local effects, which see some firms producing up to 15% less or 25% more when we remove non-salaried labor.

Table 5 shows the aggregate variation caused by the removal of non-salaried labor. As we argued, the two countervailing effects on labor demand cancel each other out, so that changes in aggregate wages, profits or output after the removal of non-salaried labor are very small. Unemployment does not change much either, increasing by only 0.005 percent in the fully flexible case. Intuitively, the removal of non-salaried workers causes a large effect on a few firms, but most firms are unaffected. Hence, these local effects are diluted when computing aggregate quantities. This suggests that the presence of informal labor does not mitigate the effects of size-dependent policies on the economy.

Figure 10 is analogous to Figure 8, but compares a regulated economy with non-salaried labor with a regulated economy without non-salaried labor. The first thing to note is that the magnitude of the variations is a lot smaller than in Figure 8. On the other hand, there are no clear winners from the removal of non-salaried labor. Any positive variation in profit or expected wages is very small, but medium-sized firms are clearly worse-off.

7.3 Changing the value of the threshold

While it is clear that removing distortionary regulations would be the first-best policy option, in practice this may not be feasible to implement for political or other reasons. With that in mind, we consider a more subtle approach by examining how the economy would respond to changes in the value of the threshold of the size-dependent policies. Note that removing all size-dependent regulations is equivalent to setting the threshold to infinity. One the other extreme, setting the threshold to 0 removes the "size-dependent" part of the policies in place, potentially reducing misallocation but increasing other distortions. In what follows, we compare the policies currently in place with similar ones, by varying the value of the threshold between 2 (the size of the smallest firm in the model) and 100.

Figure 11 shows how the economy responds to variations in the value of the threshold. The top right panel 11a shows how wages, profits and average firm size would respond to changes in the threshold under the model with flexible wages, while the top left panel 11b considers the extended model with partial wage rigidity and firm splitting.[19] A higher threshold would increase all three measures, but not enough to fully neutralize the negative effect of the policy.[20] For example, setting the threshold at 50 salaried employees instead of 20 salaried employees would lead to an increase in profits of about 0.5 percent, and average firm size would increase by about 0.2 percent.

The middle panels 11c and 11d show how the share of informal, non-salaried workers and the unemployment rate vary with different values of the threshold. The unemployment rate, shown only in panel 11d because there is no unemployment when wages are flexible, decreases with the size of the threshold. It does so for the same reason that wages, profits (panels 11a and 11b) and output (in panel 11f) increase: a larger threshold means that fewer firms are affected by the regulation, which means that those firms have lower labor costs and therefore hire more employees. For the case of profits, a higher threshold also means that fewer firms are profit-sharing, which mechanically increases aggregate profits.

The share of informal employees in the economy increases with the threshold. This is driven by the fact that, as the threshold increases, the firms hiring non-salaried employees (i.e. the

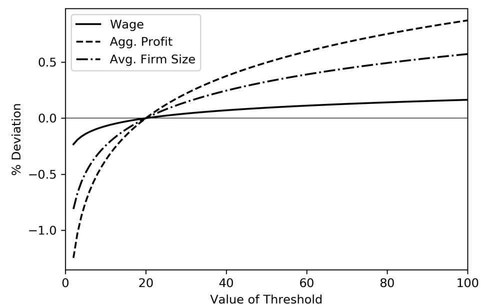
[19] As discussed in the appendix F, computing this counterfactual experiment requires a normalization of the wage rigidity equation, since we are considering infinitesimal changes.

[20] Even when the threshold is set at 100 employees, the percentage gains estimated in Figure 11 are still considerably smaller than the numbers presented in Table 4. This is because firms with more than 100 employees still represent a very large share of employment in Peru, and these firms are still being regulated.

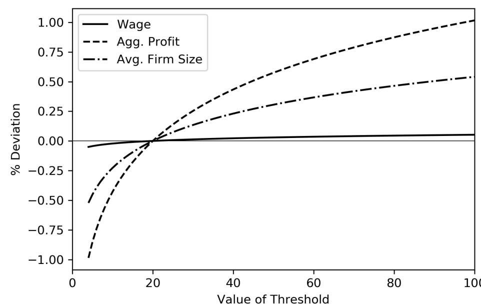
Figure 11: Adopting a different policy threshold
(% change relative to existing threshold)

Wages, Profits and Firm Size

(a) Baseline Model

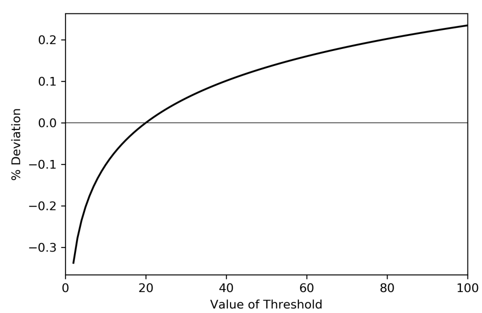


(b) With Firm Splitting and Rigid Wages

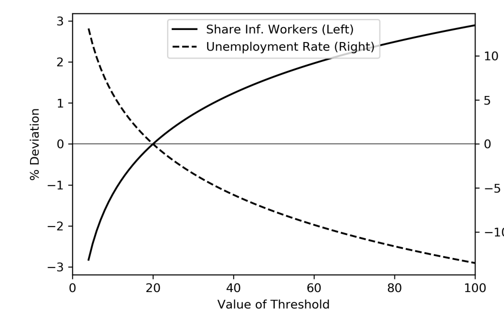


Informality and Unemployment

(c) Baseline Model

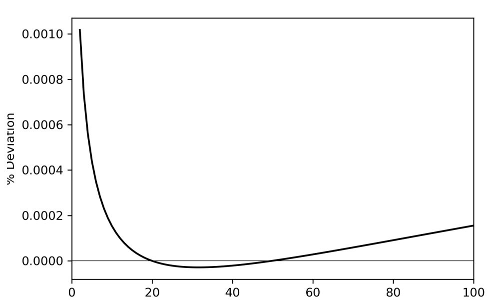


(d) With Firm Splitting and Rigid Wages

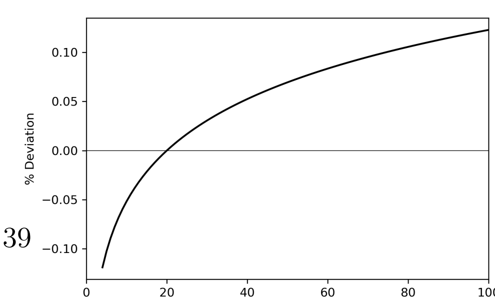


Output

(e) Baseline Model



(f) With Firm Splitting and Rigid Wages



"constrained" firms in figure 4) are bigger. Since those firms are substituting salaried employees for non-salaried ones, it necessarily follows that the share of informal workers increases.[21,22]

Finally, the bottom panels 11e and 11f show the response of aggregate output. There are two striking features in those two panels. First, the magnitude of the changes in panel 11e are a lot smaller than the magnitudes in panel 11f, as we would expect from table 4. This highlights how labor market rigidities can impede the ability of the economy to adjust to the size-dependent regulations. Second, the shape of the curves is very different, a product of the different offsetting forces at play.

On one hand, a higher threshold implies that fewer firms are affected by the regulations, which would lessen their negative impact on output. On the other hand, a higher threshold also means that affected firms — which are those closer to threshold — are now much bigger. This would worsen the aggregate impact of the regulations, since these larger firms are responsible for hiring more workers and producing more output. In addition, it can be shown that the range of most affected firms increases with the threshold value, which again aggravates the impact of the regulations in place (i.e. the difference $\alpha_r - \alpha_c$ increases with N). Meanwhile, by decreasing the threshold, more firms are affected by the size-dependent policies, which increases the impact these policies have. However, when the threshold is small, there is also less room for firms to react to those same policies, which leads to lower misallocation (i.e. there are fewer firms bunching at N salaried employees). Indeed, if the threshold is smaller than the size of the smallest firm in the economy, size-dependent policies are no longer size-dependent, as they affect all firms.

The balance between all of these effects is showcased in panel 11e, where output variations

[21] The increase in the share of informal workers is, however, not monotonic. For high enough values of the threshold, it is possible that $\alpha_r > \alpha_{\max}$, so that even the largest firms in the economy are constrained and thus hiring non-salaried workers. At that point, increasing the value of the threshold will only reduce the measure of firms who are constrained and thus the share of informal workers decreases. If we keep increasing the value of the threshold, eventually we will reach a point where $\alpha_c > \alpha_{\max}$ and all firms are smaller than the threshold (which is equivalent to removing all size-dependent policies). At this point, the share of informal workers is zero. With all of that said, the share of informal workers only starts to decrease for unrealistically high values of the threshold (around many thousands of workers), so it is not of particular relevance from a policy perspective.

[22] As mentioned previously, our model does not address all other factors behind non-salaried work in Peru, many of which are unrelated to the size-dependent policies. About 12 percent of the Peruvian workforce is non-salaried, of which 0.3 percent are hired by firms with 20 salaried employees. This suggests that 0.3 percent of the workforce is employed in a non-salaried position because of the size-dependent policies. While we do not target this moment in the estimation, the equivalent share in our model is 0.29, suggesting that the model does a good job in determining how many non-salaried jobs are caused by the size-dependent policies.

display a non-monotonic pattern. For small values of the threshold the “lower-misallocation” effect dominates, while for large values (i.e., from 60 salaried employees onwards) the fact that fewer firms are affected becomes the most important. In between those values, the negative effects of increasing the threshold cause output to fall below the current levels. Finally, it should be noted that the shape of the curve in panel 11e is extremely sensitive to the value of γ, the shape parameter of the firm-size distribution – as expected since this parameter controls how many firms would be above or below any threshold.

In panel 11f, we still have the same forces in play, but in this case a fraction of firms is able split, and so the value of the threshold is irrelevant for them. This undermines the “lower-misallocation” effect of reducing the threshold. Wages are also not allowed to fully adjust, so any indirect effect of changing the threshold becomes smaller. Hence, because a larger threshold means that fewer firms are affected by distortionary regulations, output monotonically increases with the value of the threshold.

8 Conclusion

Labor informality and weak enforcement are pervasive in many developing economies. In this paper, we examine how these structural features of the economy interact with size-dependent firm policies, and the trade-offs involved. We focus on the experience of Peru with profit sharing and other policies that are applicable only to firms with more than 20 formal workers. Using an estimated structural model, we find that, as currently designed, these policies have an overall negative effect on aggregate output. While the owners of small firms win, larger firms and workers lose. The size-dependent policies also lead to an increase in labor informality, as some firms opt to hire informal workers to avoid the regulation. If the purpose of these policies is to raise workers income, our results suggest that they make labor worse off through lower wages and higher informality and unemployment.

There are several avenues for future research. First, it would be important to obtain direct evidence of firm splitting in Peru and other economies, in particular by linking legally separate firms to the same owner. A better understanding of the costs, trade-offs and prevalence of firm splitting would be useful to improve government’s enforcement capacity, and to obtain a clearer picture of the relation between firm size and firm productivity in developing economies. Second, future work could focus on other types of size-dependent policies, including simplified tax regimes and minimum thresholds, that are ubiquitous in developing economies. Improving the design of these policies can have significant impacts on tax collection and efforts to formalize both workers and firms.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, "Identification properties of recent production function estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Alaimo, Verónica, Mariano Bosch, Melany Gualavisí, and Juan Miguel Villa, "Measuring the Cost of Salaried Labor in Latin America and the Caribbean," June 2017.

Amirapu, Amrit and Michael Gechter, "Labor Regulations and the Cost of Corruption: Evidence from the Indian Firm Size Distribution," 2017.

Bertrand, Marianne, Chang-Tai Hsieh, and Nick Tsivanidis, "Contract Labor and Firm Growth in India," 2017.

Busso, Matías, María Victoria Fazio, and Santiago Levy, "(In)Formal and (Un)Productive: The Productivity Costs of Excessive Informality in Mexico," August 2012.

Gandhi, Amit, Salvador Navarro, and David Rivers, "On the Identification of Gross Output Production Functions," 2017.

Garicano, Luis, Claire Lelarge, and John Van Reenen, "Firm Size Distortions and the Productivity Distribution : Evidence from France," *American Economic Review*, 2016, *106* (11), 3439–3479.

Gourio, François and Nicolas A. Roys, "Size-Dependent Regulations, Firm Size Distribution, and Reallocation," *Quantitative Economics*, 2014, *5* (July), 377–416.

Hsieh, Chang-Tai and Benjamin A. Olken, "The Missing "Missing Middle"," *Journal of Economic Perspectives*, 2014, *28* (3), 89–108.

_ **and Peter J. Klenow**, "The Life Cycle of Plants in India and Mexico," *The Quarterly Journal of Economics*, 2014, *129* (August), 1035–1084.

Levy, Santiago, *Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico*, Brookings Institution Press, 2008.

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

Olley, G. Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64* (6), 1263–1297.

Onji, Kazuki, "The response of fi rms to eligibility thresholds: Evidence from the Japanese value-added tax," *Journal of Public Economics*, 2009, *93* (5-6), 766–775.

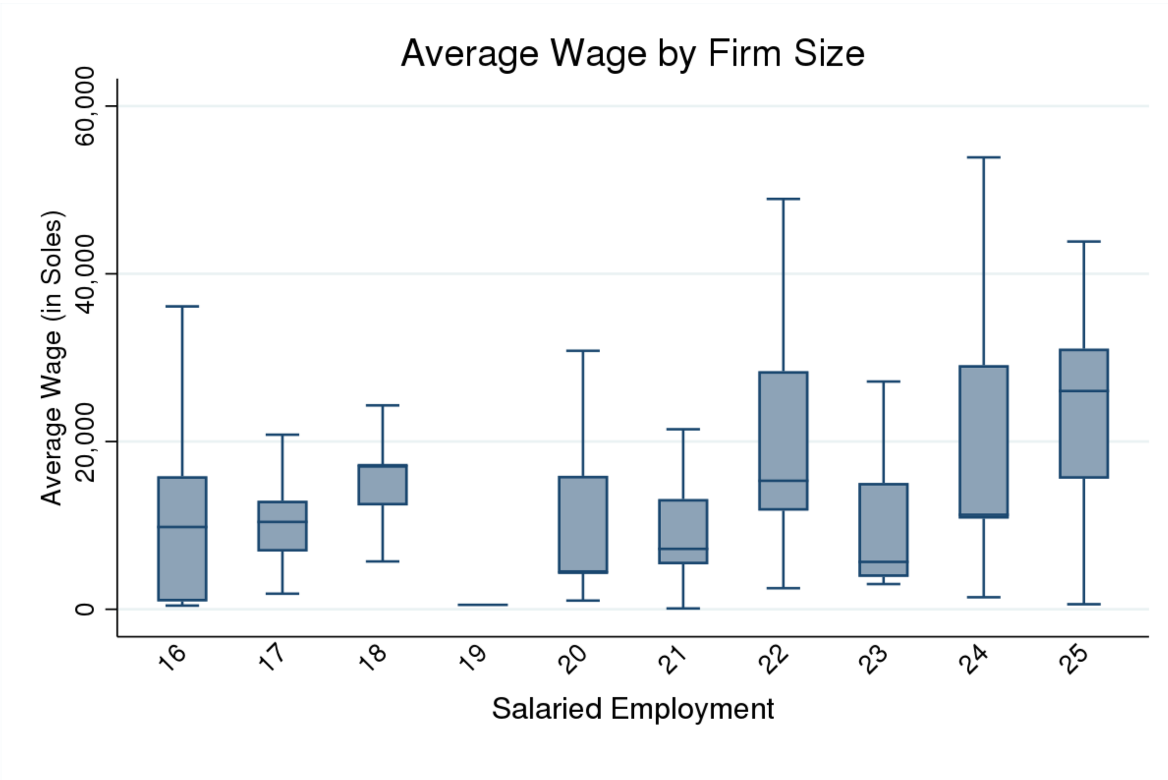
Schivardi, Fabiano and Roberto Torrini, "Identifying the effects of firing restrictions through size-contingent differences in regulation," *Labour Economics*, 2008, *15*, 482–511.

Ulyssea, Gabriel, "Firms , Informality and Development: Theory and evidence from Brazil," *American Economic Review*, 2018.

Viollaz, Mariana, "Are Labor Inspections Protecting Workers' Rights? Adding the Evidence from Size-based Labor Regulations and Fines in Peru," *International Labour Review*, 2018.

A Other tables and figures

Figure A.1: Wages around the policy threshold - ENE



Note: Average wages are computed as total personnel expenses divided by the number of workers (both salaried and non-salaried). Each box depicts the 25th, 50th and 75th percentiles of the average wage for different firms. The whiskers represent the adjacent values of the distribution (i.e. upper/lower quartile +/- 1.5 $\times$ interquartile range). The data is from the ENE 2015 firm survey, and refer to the year 2014.

B Computing salaried employees' compensations

Let $v(\cdot)$ be the worker's indirect utility function, which in this model is only a function of the worker's income. We assume that v is multiplicatively separable in the sense that $v(xy) = v(x)v(y)$.[23] Every firm is subject to an i.i.d. ξ such that the firm's realized profit is $\xi\pi(\alpha)$, where $\mathbb{E}[\xi] = 1$.[24]

Since all workers are ex-ante identical, we must have

$$\mathbb{E}\left[v\left(c(\tau_\pi, \tau_p) + \frac{\tau_\pi \xi \pi(\alpha)}{\ell(\alpha)}\right)\right] = v(w).$$

That is, the expected compensation for a worker that has a share of profits has to be equal to the wage of the worker who does not have a stake in the firm's profit. Given that firms who are actively hiring salaried workers will not hire non-salaried workers (see the discussion on the main text), we can solve the firm's profit maximization problem

$$\mathbb{E}[\xi\pi(\alpha)] = \pi(\alpha) = \max_{\ell}(1 - \tau_\pi)[\alpha g(\ell) - (1 + \tau_w + \tau_p)c(\tau_\pi, \tau_p)\ell]$$

Using $g(x) = x^\theta$, we find that

$$\frac{\pi(\alpha)}{\ell(\alpha)} = \frac{1 - \theta}{\theta}(1 + \tau_w + \tau_p)c(\tau_\pi, \tau_p).$$

Plugging this into the indifference equation above, we get

$$v(c(\tau_\pi, \tau_p))\mathbb{E}\left[v\left(1 + \xi\tau_\pi \frac{1 - \theta}{\theta}(1 + \tau_w + \tau_p)\right)\right] = v(w),$$

where we have used the multiplicative separability of $v(\cdot)$ to factor out $v(c(\tau_\pi, \tau_p))$. Define $\bar{k}$ as

$$\frac{1}{v(1/\bar{k})} = \mathbb{E}\left[v\left(1 + \xi\tau_\pi \frac{1 - \theta}{\theta}(1 + \tau_w + \tau_p)\right)\right].$$

Then

$$v(c(\tau_\pi, \tau_p)) = v(1/\bar{k})v(w) = v(w/\bar{k})$$

[23] This assumption is true for many commonly used utility functions, e.g. CRRA.

[24] The form of the shock is of little consequence for our result. An additive shock, for example $\pi(\alpha) + \xi$, would lead to the same conclusion.

which immediately implies $c(\tau_\pi, \tau_p) = w/\bar{k}$. Finally, define $T = \frac{(1+\tau_w+\tau_p)}{\bar{k}}$, we have that

$$Tw = \frac{(1+\tau_w+\tau_p)}{\bar{k}} w = (1+\tau_w+\tau_p) c(\tau_\pi, \tau_p),$$

as desired. Also note from the definition of $\bar{k}$ that $v(1/\bar{k}) \leq 1$[25], which means that $v(c(\tau_\pi, \tau_p)) \leq v(w)$ and therefore $c(\tau_\pi, \tau_p) \leq w$ (i.e. firms are paying their employees a smaller wage to compensate for the profit sharing rule). On the other hand, the fact that we don't observe small firms sharing profits as well — in fact they go out of their way to avoid sharing profits — suggests that $T \geq 1 + \tau_w$. Put differently, the reduction in wage made possible by the profit sharing rule is not large enough to undo the rise in labor costs caused by the health and unionizing rules that apply to larger firms.

C Estimating the elasticity of labor

C.1 Estimate θ

As pointed out in recent papers by Ackerberg et al. (2015) and Gandhi et al. (2017), the proxy variable method pioneered by Olley and Pakes (1996) may not identify the parameters in the production function. With that in mind, we propose calibrating θ to match the mean labor cost share of firms. Note that, from the first order conditions of the firm's problem, we have

$$\alpha\theta(\ell + \phi n)^{\theta-1} = w(1+\tau).$$

Simple algebra leads us to the equality

$$\theta = \frac{w(1+\tau)(\ell+\phi n)}{\alpha(\ell+\phi n)^\theta}$$

[25] From the definition of $\bar{k}$, we have that,

$$\frac{1}{v(1/\bar{k})} = \mathbb{E}\left[v\left(1 + \xi\tau_\pi \frac{1-\theta}{\theta}(1+\tau_w+\tau_p)\right)\right] \geq v(1),$$

because $\xi \geq 0$ and v is an increasing function (as it is an indirect utility function). It follows that $\frac{1}{v(1/\bar{k})} \geq v(1) \implies v(1/\bar{k})v(1) \leq 1 \implies v(1/\bar{k}) \leq 1$.

Table A.1: Distribution of the firm's labor cost share.

	Mean	Std. Dev.	p10	p25	p50	p75	p90	N
Labor cost share	0.67	0.25	0.31	0.47	0.71	0.91	0.93	1,233

For firms who do not hire non-salaried labor, this equality reduces to

$$\theta = \frac{w(1+\tau)\ell}{\alpha \ell^{\theta}} \equiv \frac{\text{wage bill}}{\text{sales}}.$$

The value of the right-hand size of the expression above is observable in our data. Using the ENE dataset, which describes a firm's expenditure in detail, we define labor expenditure as the sum of expenditures on own personnel and on services provided by third party firms. We can simply divide the labor expenditure by the value of sales to get a value for our right-hand size variable. However, this leads to many instances in which the labor cost share would be virtually zero or a lot larger than 1. Because of this, we take a slightly different approach and use the total cost of a firm as the denominator in the expression above.[26].

The distribution of the labor cost share (restricted to firms who do not hire non-salaried employees) is shown in table A.1. Note that the mean and median of the distribution are 0.67 and 0.71, which are very close to the values we estimated using the regressions in section 6.2.1.

C.2 Estimate ϕ

To estimate ϕ, we use 3 different methods. The first method is described in section 6.2.1. We can also estimate ϕ using the most direct way and run a non-linear regression of log(sales) on $\theta \log(\ell + \phi n)$ and the same controls as in equation 7. We do not include fixed effects, since that could generate numerical issues by making the parameter space too big.

Note that we can write

$$\theta \log(\ell + \phi n) = \theta \log(\ell) + \theta \log\left(1 + \phi \frac{n}{\ell}\right) \approx \theta \log(\ell) + \theta \phi \frac{n}{\ell}.$$

This suggests a third specification that regresses log(sales) on $\log(\ell)$ and the labor ratio n/ℓ (plus all the remaining controls in Equation (7)). The ratio between the coefficients on the

[26]Besides the own personnel expenditures and third party expenditures, our total expenditure variable includes taxes, materials and other inputs and "other expenditures".

Table A.2: Regressions of log-sales on level of inputs and controls

	(1: non-linear)	(2)	(3)
$\hat{\theta}$	0.745***		
	(0.08)		
$\hat{\phi}$	0.846		
	(0.533)		
log salaried		0.729***	
		(0.064)	
labor ratio		0.499**	
		(0.235)	
# Salaried			0.074***
			(0.019)
# Non-Salaried			0.062***
			(0.005)
Implied ϕ	0.846	0.685	0.831***
	(0.533)	(0.488)	(0.107)
Controls	No sector or firm type FE	Yes	Yes
Sample	Full	labor ratio ≤ 2	# Salaried < 20
Observations	2214	2147	2013
R-squared	0.99	0.69	0.66

Note: All specifications control for average worker education, 4-digit industry sector and firm type (e.g. personal, LLC, SA). Standard errors are robust to heteroskedacity. * 0.10, ** 0.05, *** 0.01.

labor ratio and salaried labor would give us ϕ. The caveat of running this regression is that the approximation $\log(1 + x) \approx x$ is not appropriate for large values of x, and therefore we restrict our sample to $n/\ell <= 2$.

The estimation results are shown in Table A.2. The standard errors for the implied ϕ are calculated using the delta-method. Columns (1) and (2) also contain estimated values for $\hat{\theta}$, and are consistent with the values found in Table 1. Column (3) repeats the results described in section 6.2.1. The estimated values for $\hat{\phi}$ are roughly consistent with each other in columns (1), (2) and (3), but the standard errors are large for the first two columns. Although not ideal, this is not totally unexpected, as data on non-salaried labor is bound to have a great deal of measurement error and is not available for all firms in our data.

D Estimating the firm size distribution

D.1 Derivation

Given the salaried labor demand specified in equation 2 and the functional forms $g(\ell) = \ell^{\theta}$ and $\psi(\alpha) = C\alpha^{-\beta}$, we can use the change in variables formula to find the density of salaried employment:

$$\zeta(\ell(\alpha)) = \frac{C\alpha(\ell)^{-\beta}}{P}\alpha'(\ell),$$

where $\alpha(\ell)$ is the inverse demand function for salaried labor and P is the measure of managers in the economy (necessary for $\zeta(\ell)$ to integrate to 1). Then

$$\zeta(\ell) = \begin{cases} \frac{C}{P}\left[\frac{(1+\tau_w)w}{\theta}\ell^{1-\theta}\right]^{-\beta}\frac{(1+\tau_w)w}{\theta}(1-\theta)\ell^{-\theta}, & \ell_{\min} \le \ell < N \\ \int_{\alpha_c}^{\alpha_r}\frac{C}{P}\alpha^{-\beta}d\alpha, & \ell = N \\ 0, & N < \ell < \ell_r \\ \frac{C}{P}\left[\frac{Tw}{\theta}\ell^{1-\theta}\right]^{-\beta}\frac{Tw}{\theta}(1-\theta)\ell^{-\theta}, & \ell_r \le \ell_{\max} \end{cases}$$

We can compute the integral in this expression by using the substitution method, where $\alpha = \ell^{1-\theta}\frac{(1+\tau_w)w}{\theta}$:

$$\begin{aligned} \int_{\alpha_c}^{\alpha_r}\frac{C}{P}\alpha^{-\beta}d\alpha &= \frac{C}{P}\int_{\left(\frac{\alpha_c\theta}{(1+\tau_w)w}\right)^{\frac{1}{1-\theta}}}^{\left(\frac{\alpha_r\theta}{(1+\tau_w)w}\right)^{\frac{1}{1-\theta}}}\left[\ell^{1-\theta}\frac{(1+\tau_w)w}{\theta}\right]^{-\beta}(1-\theta)\ell^{-\theta}\frac{(1+\tau_w)w}{\theta}d\ell \\ &= \frac{C}{P}\left(\frac{(1+\tau_w)w}{\theta}\right)^{1-\beta}(1-\theta)\int_{N}^{\left(\frac{T}{1+\tau_w}\right)^{\frac{1}{1-\theta}}\ell_r}\ell^{-\beta(1-\theta)-\theta}d\ell \\ &= \frac{C}{P}\left(\frac{(1+\tau_w)w}{\theta}\right)^{\frac{1-\gamma}{1-\theta}}\frac{1-\theta}{1-\gamma}\ell^{1-\gamma}\Big|_{N}^{\left(\frac{T}{1+\tau_w}\right)^{\frac{1}{1-\theta}}\ell_r} \end{aligned}$$

where we have used $\beta(1-\theta)+\theta=\gamma$ in the last equality. It follows that

$$\zeta(\ell)=\begin{cases}\frac{C}{P}\left(\frac{\theta}{(1+\tau_w)w}\right)^{\frac{\gamma-1}{1-\theta}}(1-\theta)\ell^{-\gamma}, & \ell_{\min}\leq\ell<N\\ \frac{C}{P}\left(\frac{\theta}{w}\right)^{\frac{\gamma-1}{1-\theta}}\left(\frac{1-\theta}{\gamma-1}\right)\left[(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}N^{1-\gamma}-T^{\frac{1-\gamma}{1-\theta}}\ell_r^{1-\gamma}\right], & \ell=N\\ 0, & N<\ell<\ell_r\\ \frac{C}{P}\left(\frac{\theta}{Tw}\right)^{\frac{\gamma-1}{1-\theta}}(1-\theta)\ell^{-\gamma}, & \ell_r\leq\ell\leq\ell_{\max}\end{cases}$$

as shown in the main text. For ease of notation, define $K=\frac{C}{P}\left(\frac{\theta}{w}\right)^{\frac{\gamma-1}{1-\theta}}\left(\frac{1-\theta}{\gamma-1}\right)$. Since $\zeta(\ell)$ is a density, we have that

$$\int_{\ell_{\min}}^{\ell_{\max}}\zeta(\ell)d\ell=1.$$

This imposes the restriction

$$\frac{1}{K}=(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}\ell_{\min}^{1-\gamma}-T^{\frac{1-\gamma}{1-\theta}}\ell_{\max}^{1-\gamma}$$

D.2 Estimation

As mentioned in the main text, to estimate the firm size distribution, we follow the method proposed by Garicano et al. (2016). This consists of assuming that we only observe labor with measurement error:

$$\ell(\alpha,\varepsilon)=e^{\varepsilon}\ell(\alpha),$$

where $\varepsilon\sim\mathcal{N}(0,\sigma)$. The conditional distribution of $\ell(\alpha,\varepsilon)$ is given by

$$\mathbb{P}(\ell(\alpha,\varepsilon)<x|\varepsilon)=\mathbb{P}(\ell(\alpha)<e^{-\varepsilon}x|\varepsilon)=\int_{\ell_{\min}}^{e^{-\varepsilon}x}\zeta(y)dy$$

Thus

$$\mathbb{P}(\ell(\alpha,\varepsilon)<x|\varepsilon)=\begin{cases}\underbrace{\frac{C}{P}\left(\frac{\theta}{(1+\tau_w)w}\right)^{\frac{\gamma-1}{1-\theta}}(1-\theta)\int_{\ell_{\min}}^{e^{-\varepsilon}x}y^{-\gamma}dy,}_{E1} & \ell_{\min}\le e^{-\varepsilon}x<N\\ \underbrace{E1|_{e^{-\varepsilon}x=N}+\frac{C}{P}\left(\frac{\theta}{w}\right)^{\frac{\gamma-1}{1-\theta}}\left(\frac{1-\theta}{\gamma-1}\right)\left[(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}N^{1-\gamma}-T^{\frac{1-\gamma}{1-\theta}}\ell_r^{1-\gamma}\right],}_{E2} & N\le e^{-\varepsilon}x<\ell_r\\ E2+\frac{C}{P}\left(\frac{\theta}{Tw}\right)^{\frac{\gamma-1}{1-\theta}}(1-\theta)\int_{\ell_r}^{e^{-\varepsilon}x}y^{-\gamma}dy, & \ell_r\le e^{-\varepsilon}x\le\ell_{\max}\end{cases}$$

Recall that we have defined $K=\frac{C}{P}\left(\frac{\theta}{w}\right)^{\frac{\gamma-1}{1-\theta}}\left(\frac{1-\theta}{\gamma-1}\right)$. Then

$$\mathbb{P}(\ell(\alpha,\varepsilon)<x|\varepsilon)=\begin{cases}K(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}\left[\ell_{\min}^{1-\gamma}-e^{\varepsilon(\gamma-1)}x^{1-\gamma}\right], & \log\left(\frac{x}{N}\right)<\varepsilon\le\log\left(\frac{x}{\ell_{\min}}\right)\\ K\left[(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}\ell_{\min}^{1-\gamma}-T^{\frac{1-\gamma}{1-\theta}}\ell_r^{1-\gamma}\right], & \log\left(\frac{x}{\ell_r}\right)<\varepsilon\le\log\left(\frac{x}{N}\right)\\ K\left[(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}\ell_{\min}^{1-\gamma}-e^{\varepsilon(\gamma-1)}T^{\frac{1-\gamma}{1-\theta}}x^{1-\gamma}\right], & \log\left(\frac{x}{\ell_{\max}}\right)\le\varepsilon\le\log\left(\frac{x}{\ell_r}\right)\end{cases}$$

Now we integrate the measurement error out,

$$\mathbb{P}(\ell(\alpha,\varepsilon))=\int_{-\infty}^{\infty}\mathbb{P}(\ell(\alpha,\varepsilon)<x|\varepsilon)\varphi\left(\frac{\varepsilon}{\sigma}\right)d\varepsilon.$$

To compute this integral, first note that

$$\int_{\log(x/a)}^{\log(x/b)}e^{\varepsilon(\gamma-1)}\varphi\left(\frac{\varepsilon}{\sigma}\right)d\varepsilon\propto\int_{\log(x/a)}^{\log(x/b)}e^{\varepsilon(\gamma-1)}e^{-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2}d\varepsilon=\int_{\log(x/a)}^{\log(x/b)}e^{\frac{\sigma^2(\gamma-1)^2}{2}-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}-\sigma(\gamma-1)\right)^2}d\varepsilon,$$

therefore

$$\int_{\log(x/a)}^{\log(x/b)}e^{\varepsilon(\gamma-1)}\varphi\left(\frac{\varepsilon}{\sigma}\right)d\varepsilon=e^{\frac{\sigma^2(\gamma-1)^2}{2}}\left\{\Phi\left(\frac{\log(x/b)}{\sigma}-\sigma(\gamma-1)\right)-\Phi\left(\frac{\log(x/a)}{\sigma}-\sigma(\gamma-1)\right)\right\}$$

where φ is the standard normal density and Φ is the standard normal distribution.

Using the result above, it is straightforward to show that

$$
\begin{aligned}
\mathbb{P}(\ell(\alpha,\varepsilon)<x) = \quad & K(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}\ell_{\min}^{1-\gamma}\left\{\Phi\left(\frac{\log(x/\ell_{\min})}{\sigma}\right)-\Phi\left(\frac{\log(x/\ell_{\max})}{\sigma}\right)\right\} \\
& -KT^{\frac{1-\gamma}{1-\theta}}\ell_r^{1-\gamma}\left\{\Phi\left(\frac{\log(x/N)}{\sigma}\right)-\Phi\left(\frac{\log(x/\ell_r)}{\sigma}\right)\right\} \\
& -K(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}x^{1-\gamma}e^{\frac{1}{2}\sigma^2(\gamma-1)^2}\left\{\Phi\left(\frac{\log(x/\ell_{\min})}{\sigma}-\sigma(\gamma-1)\right)-\Phi\left(\frac{\log(x/N)}{\sigma}-\sigma(\gamma-1)\right)\right\} \\
& -KT^{\frac{1-\gamma}{1-\theta}}x^{1-\gamma}e^{\frac{1}{2}\sigma^2(\gamma-1)^2}\left\{\Phi\left(\frac{\log(x/\ell_r)}{\sigma}-\sigma(\gamma-1)\right)-\Phi\left(\frac{\log(x/\ell_{\max})}{\sigma}-\sigma(\gamma-1)\right)\right\}.
\end{aligned}
$$

Finally, since we are interested in the density of $\ell(\alpha,\varepsilon)$, we compute

$$\omega(x)=\frac{\partial\mathbb{P}(\ell(\alpha,\varepsilon)<x)}{\partial x}.$$

Taking the derivative, we find

$$
\begin{aligned}
\omega(x) = \quad & K(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}\ell_{\min}^{1-\gamma}\left\{\varphi\left(\frac{\log(x/\ell_{\min})}{\sigma}\right)-\varphi\left(\frac{\log(x/\ell_{\max})}{\sigma}\right)\right\}\frac{1}{\sigma x} \\
& -KT^{\frac{1-\gamma}{1-\theta}}\ell_r^{1-\gamma}\left\{\varphi\left(\frac{\log(x/N)}{\sigma}\right)-\varphi\left(\frac{\log(x/\ell_r)}{\sigma}\right)\right\}\frac{1}{\sigma x} \\
& -K(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}x^{1-\gamma}e^{\frac{1}{2}\sigma^2(\gamma-1)^2}\left\{\varphi\left(\frac{\log(x/\ell_{\min})}{\sigma}-\sigma(\gamma-1)\right)-\varphi\left(\frac{\log(x/N)}{\sigma}-\sigma(\gamma-1)\right)\right\}\frac{1}{\sigma x} \\
& -KT^{\frac{1-\gamma}{1-\theta}}x^{1-\gamma}e^{\frac{1}{2}\sigma^2(\gamma-1)^2}\left\{\varphi\left(\frac{\log(x/\ell_r)}{\sigma}-\sigma(\gamma-1)\right)-\varphi\left(\frac{\log(x/\ell_{\max})}{\sigma}-\sigma(\gamma-1)\right)\right\}\frac{1}{\sigma x} \\
& +(\gamma-1)K(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}x^{-\gamma}e^{\frac{1}{2}\sigma^2(\gamma-1)^2}\left\{\Phi\left(\frac{\log(x/\ell_{\min})}{\sigma}-\sigma(\gamma-1)\right)-\Phi\left(\frac{\log(x/N)}{\sigma}-\sigma(\gamma-1)\right)\right\} \\
& +(\gamma-1)KT^{\frac{1-\gamma}{1-\theta}}x^{-\gamma}e^{\frac{1}{2}\sigma^2(\gamma-1)^2}\left\{\Phi\left(\frac{\log(x/\ell_r)}{\sigma}-\sigma(\gamma-1)\right)-\Phi\left(\frac{\log(x/\ell_{\max})}{\sigma}-\sigma(\gamma-1)\right)\right\}.
\end{aligned}
$$

Most of the terms in this expression cancel out once we note that, for any $L \neq 0$,

$$x^{1-\gamma} e^{\frac{1}{2}\sigma^2(\gamma-1)^2} \varphi\left(\frac{\log(x/L)}{\sigma} - \sigma(\gamma-1)\right) = X^{1-\gamma}\varphi\left(\frac{\log(x/L)}{\sigma}\right)$$

so that we're left with

$$\begin{aligned}
\omega(x) = \; & \frac{K}{\sigma x}\left(T^{\frac{1-\gamma}{1-\theta}}\ell_{\max}^{1-\gamma} - (1+\tau_w)^{\frac{1-\gamma}{1-\theta}}\ell_{\min}^{1-\gamma}\right)\varphi\left(\frac{x/\ell_{\max}}{\sigma}\right) \\
& + \frac{K}{\sigma x}\left((1+\tau_w)^{\frac{1-\gamma}{1-\theta}}N^{1-\gamma} - T^{\frac{1-\gamma}{1-\theta}}\ell_r^{1-\gamma}\right)\varphi\left(\frac{x/N}{\sigma}\right) \\
& + (\gamma-1)K(1+\tau_w)^{\frac{1-\gamma}{1-\theta}}x^{-\gamma}e^{\frac{1}{2}\sigma^2(\gamma-1)^2}\left\{\Phi\left(\frac{\log(x/\ell_{\min})}{\sigma} - \sigma(\gamma-1)\right) - \Phi\left(\frac{\log(x/N)}{\sigma} - \sigma(\gamma-1)\right)\right\} \\
& + (\gamma-1)KT^{\frac{1-\gamma}{1-\theta}}x^{-\gamma}e^{\frac{1}{2}\sigma^2(\gamma-1)^2}\left\{\Phi\left(\frac{\log(x/\ell_r)}{\sigma} - \sigma(\gamma-1)\right) - \Phi\left(\frac{\log(x/\ell_{\max})}{\sigma} - \sigma(\gamma-1)\right)\right\}.
\end{aligned}$$

Maximum likelihood problem Given $\omega(\cdot)$, our estimation becomes a standard maximum likelihood problem, with the exception that we require one last model constraint to hold: firms hiring salaried employees must find it optimal to do so. In other words, when firms "return" to hiring salaried employees and pay the extra costs imposed by the policies in place, their profits must be at least as high as they would have been if the same firm hired non-salaried employees and avoided the regulation. As mentioned in the main text, this requires

$$(1-\tau_\pi^{\min})[\alpha_r g(\ell(\alpha_r)) - Tw\ell(\alpha_r)] \geq \alpha_r g(N + \phi n(\alpha_r)) - w(1+\tau_w)N - wn(\alpha_r)$$

where $tau_\pi^{\min} = 0.05$ is the smallest possible profit sharing requirement for any firm.

Using the functional form of g, the first order conditions on ℓ and n imply that

$$\alpha_r = \ell(\alpha_r)^{1-\theta}\frac{Tw}{\theta} \quad \text{and} \quad n(\alpha_r) = \ell(\alpha_r)T^{\frac{1}{1-\theta}}\phi^{\frac{\theta}{1-\theta}} - \frac{N}{\phi}.$$

Plugging those into the expression above, we find (after some algebra) that

$$(1-\tau_\pi^{\min})\ell_r T\left(\frac{1-\theta}{\theta}\right) \geq \left(\frac{1-\theta}{\theta}\right)\left[\ell_r T^{\frac{1}{1-\theta}}\phi^{\frac{\theta}{1-\theta}} - \frac{N}{\phi}\right] + \frac{N}{\theta\phi} - (1+\tau_w)N$$

or

$$\ell_r \geq \frac{N\left(\frac{1}{\phi} - (1+\tau_w)\right)}{\left(\frac{1-\theta}{\theta}\right)\left[T(1-\tau_\pi^{\min}) - T^{\frac{1}{1-\theta}}\phi^{\frac{\theta}{1-\theta}}\right]} \tag{8}$$

The maximum likelihood problem we solve is thus

$$\max_{\gamma,\sigma,\ell_r,T,\ell_{\max}} \log \mathcal{L} = \sum_j \log \omega(\ell_j) \quad \text{subject to (8).}$$

D.3 Remaining parameters in the model

The list of remaining parameters in the model is $\alpha_{\max}, \alpha_a, \alpha_c, \underline{\alpha}, \alpha_{\min}, w, C, P$ and τ_π. As mentioned, we normalize $\alpha_{\max} = 1$, which immediately gives us the equilibrium wages through the FOC of the managers with productivity $\alpha_{\max}$:

$$\ell_{\max}^{1-\theta} = \frac{\alpha_{\max}\theta}{Tw} \Longrightarrow w = \frac{\alpha_{\max}\theta}{T\ell_{\max}^{1-\theta}}.$$

Once we have w, we can find all of the thresholds:

$$\alpha_c = \frac{w(1+\tau_w)}{\theta}N^{1-\theta}$$

$$\alpha_a = \alpha_c \frac{1}{(1+\tau_w)\phi}$$

$$\alpha_r = \frac{wT}{\theta}\ell_r^{1-\theta}$$

$\underline{\alpha}$ can be found by the condition

$$\underline{\alpha}[\ell(\underline{\alpha})]^\theta - (1+\tau_w)w\ell(\underline{\alpha}) = w$$

which yields

$$\underline{\alpha} = \frac{w(1+\tau_w)^\theta}{\theta^\theta(1-\theta)^{1-\theta}}$$

τ_π is also found using the restriction

$$(1-\tau_\pi)[\alpha_r g(\ell(\alpha_r)) - Tw\ell(\alpha_r)] = \alpha_r g(N+\phi n(\alpha_r)) - w(1+\tau_w)N - wn(\alpha_r),$$

which gives $\tau_\pi = 5.6\%$.

To determine $\alpha_{\min}$, we use the market clearing condition

$$\int_{\alpha_{\min}}^{\underline{\alpha}} C\alpha^{-\beta}d\alpha = \int_{\underline{\alpha}}^{\alpha_c} \ell(\alpha)C\alpha^{-\beta}d\alpha + N\int_{\alpha_c}^{\alpha_a} C\alpha^{-\beta}d\alpha + \int_{\alpha_a}^{\alpha_r}(N+n(\alpha))C\alpha^{-\beta}d\alpha + \int_{\alpha_r}^{\alpha_{\max}} \ell(\alpha)C\alpha^{-\beta}d\alpha.$$

Note that all of the C's are canceled off and we are left with an expression that a function of estimated parameters, the remaining thresholds and the wage. Skipping through most of the algebra, the market clearing condition boils down to

$$\begin{aligned}
\left(\frac{1-\theta}{\gamma-1}\right)\left[\alpha_{\min}^{\frac{1-\gamma}{1-\theta}} - \underline{\alpha}^{\frac{1-\gamma}{1-\theta}}\right] &= \left(\frac{\theta}{w(1+\tau_w)}\right)^{\frac{1}{1-\theta}} \frac{1-\theta}{2-\gamma}\left[\alpha_c^{\frac{2-\gamma}{1-\theta}} - \underline{\alpha}^{\frac{1-\gamma}{2-\theta}}\right] + N\left(\frac{1-\theta}{\gamma-1}\right)\left[\alpha_c^{\frac{1-\gamma}{1-\theta}} - \alpha_a^{\frac{1-\gamma}{1-\theta}}\right] \\
&+ \left(\frac{\theta}{w}\right)^{\frac{1}{1-\theta}} \phi^{\frac{\theta}{1-\theta}} \frac{1-\theta}{2-\gamma}\left[\alpha_r^{\frac{2-\gamma}{1-\theta}} - \alpha_a^{\frac{2-\gamma}{1-\theta}}\right] - N\left(\frac{1-\phi}{\phi}\right)\left(\frac{1-\theta}{\gamma-1}\right)\left[\alpha_a^{\frac{1-\gamma}{1-\theta}} - \alpha_r^{\frac{1-\gamma}{1-\theta}}\right] \\
&+ \left(\frac{\theta}{Tw}\right)^{\frac{1}{1-\theta}} \frac{1-\theta}{2-\gamma}\left[\alpha_{\max}^{\frac{2-\gamma}{1-\theta}} - \alpha_r^{\frac{2-\gamma}{1-\theta}}\right]
\end{aligned}$$

where we have used that $\gamma = \beta(1-\theta)+\theta$. Note that this is the labor market clearing condition for the model where firms are not allowed to split. When firms do split, the market clearing condition is slightly different, since firms that have split don't need to pay the labor costs imposed by the size dependent policies and therefore demand more labor.

Having a value for $\alpha_{\min}$, we can compute C using the fact that $\psi(\alpha)$ is a density

$$\int_{\alpha_{\min}}^{\alpha_{\max}} C\alpha^{-\beta}d\alpha = 1$$

so that

$$C = \frac{1-\gamma}{1-\theta}\left[\alpha_{\max}^{\frac{1-\gamma}{1-\theta}} - \alpha_{\min}^{\frac{1-\gamma}{1-\theta}}\right]^{-1}.$$

Finally,

$$P = \int_{\underline{\alpha}}^{\alpha_{\max}} C\alpha^{-\beta} d\alpha = C\frac{1-\theta}{\gamma-1}\left[\left(\frac{w(1+\tau_w)^{\theta}}{\theta^{\theta}(1-\theta)^{1-\theta}}\right)^{\frac{1-\gamma}{1-\theta}} - \alpha_{\max}^{\frac{1-\gamma}{1-\theta}}\right]$$

E Model with firm splitting

E.1 Derivation

Our starting point for deriving the firm size distribution when firms are able to split is the density in equation 6. The main effect of firm splitting on the density $\zeta(\ell)$ is a redistribution of mass from firms with more than ℓ_r salaried employees to smaller firms. In particular, since a fraction δ of large firms will split, $\zeta^s(\ell) = (1-\delta)\zeta(\ell)$ for $\ell > \ell_r$, where ζ^s is the firm size density when firms split. Since there are no firms with $\ell \in (N, \ell_r)$ salaried employees and no firms who split into units with less than $N/2$ employees, $\zeta^s(\ell) = \zeta(\ell)$ for $\ell \in [\ell_{\min}, N/2] \cup (N, \ell_r)$.

Thus, the density of firms will only increase for firms with $\ell \in (N/2, N]$ salaried employees. However, the increase is asymmetric across firms. As firms grow, they will start to split into sub-firms whose size is bounded below by larger numbers. For example, the number of firms with 12 employees increases because firms with 24 employees will split into two firm of 12. However, no larger firm will optimally split into firms of 12 employees each. A firm with 36 employees could potentially split into 3 sub-firms of 12 employees each, but by assumption 2, it will instead break into 2 sub-firms of 18 employees.

As a general rule, for each number ℓ between $N/2$ and N, the size of the biggest firm that splits into units of size ℓ is given by ℓJ_ℓ (which means this firm splits into J_ℓ pieces), where

$$J_\ell = \max\left\{\left\lceil \frac{\ell}{N-\ell}\right\rceil, 1\right\} \quad \text{and} \quad J_N = \left\lfloor \frac{\ell_{\max}}{N} \right\rfloor \tag{9}$$

To see why this is true, we follow a simple logic derived from the assumptions we made about how firms split. Namely, if a firm splits into x pieces of size ℓ, it does so because if it can't split into a smaller number of pieces; i.e. if it split into fewer pieces, each piece would be larger than N.

Let $x\ell$ be the size of the largest firm that splits into units of size ℓ. The next candidate firm that could split into units of size ℓ has total size $(x+1)\ell$. Since by definition this firm does not split into units of size ℓ, we know that $\frac{(x+1)\ell}{x} \leq N$ – that is, if the firm split into x pieces instead of $(x+1)$ pieces, each piece would still be smaller than N.

Example: An example might help grasp the logic. Consider all of the possible candidate firms who would split into units of 12 salaried employees each:

total number of employees	12	24	36	48	60	72	...
splits into	1×12	2×12	2×18	3×16	3×20	4×18	...

A firm with 24 employees will divide into 2 units of 12 employees each because it cannot avoid the regulation costs by not splitting. A firm with 36 employees could also split into 3 units of 12 employees, but it also has the option of splitting into 2 units of 18 – i.e. $\frac{3\times 12}{2} = 18 \leq 20$. The same is true for a firm with 48 employees: instead of splitting into 4 units of 12, it can split into 3 units of 16, and so on.

Going back to the general case, it follows that

$$J_\ell = \arg\min\{x : (x+1)\ell \leq Nx\} = \arg\min\left\{x : x \geq \frac{\ell}{N-\ell}\right\}$$

Since J_ℓ is an integer, it will be given by 9. Note that J_ℓ is increasing on ℓ, which means that the density of firms will increase more the closer they are to having N employees. In particular, the density of firms with n employees, $\ell_{\text{min}} \leq n < N$ will now be

$$\zeta^s(n) = K(\gamma-1)(1+\tau_w)^{\frac{1-\gamma}{1-\theta}} n^{-\gamma} + \delta \sum_{j=2}^{J_n} K(\gamma-1)(1+\tau_w)^{\frac{1-\gamma}{1-\theta}} j(jn)^{-\gamma}$$

where the second term (in red) is the sum of the density of all firms that split into units of size n (which have themselves size jn and break into j different firms). Note also that because those firms are able to split, they do not pay the extra labor costs, T) and recall that $K = \frac{C}{P}\left(\frac{\theta}{w}\right)^{\frac{\gamma-1}{1-\theta}}\left(\frac{1-\theta}{\gamma-1}\right)$

We can rewrite $\zeta^s(n)$ as

$$\zeta^s(n) = K(\gamma-1)(1+\tau_w)^{\frac{1-\gamma}{1-\theta}} n^{-\gamma}\left[(1-\delta) + \delta \sum_{j=1}^{J_n} j^{1-\gamma}\right]$$

When the firm has exactly N employees, J_ℓ is defined as

$$J_N = \left\lfloor \frac{\ell_{\max}}{N} \right\rfloor.$$

Following the reasoning above, we get

$$\zeta^s(N) = (1-\delta)K\left[(1+\tau_w)^{\frac{1-\gamma}{1-\theta}} N^{-\gamma} - T^{\frac{1-\gamma}{1-\theta}} \ell_r^{-\gamma}\right] + \delta K(\gamma-1)(1+\tau_w)^{\frac{1-\gamma}{1-\theta}} N^{-\gamma} \sum_{j=1}^{J_N} j^{1-\gamma}.$$

Putting everything together,

$$\zeta^s(\ell) = \begin{cases} 0, & \ell < \ell_{\min} \\ K(\gamma-1)(1+\tau_w)^{\frac{1-\gamma}{1-\theta}} \ell^{-\gamma} \left[(1-\delta) + \delta \sum_{j=1}^{J_\ell} j^{1-\gamma}\right], & \ell_{\min} \le \ell < N \\ (1-\delta)K\left[(1+\tau_w)^{\frac{1-\gamma}{1-\theta}} N^{-\gamma} - T^{\frac{1-\gamma}{1-\theta}} \ell_r^{-\gamma}\right] + \delta K(\gamma-1)(1+\tau_w)^{\frac{1-\gamma}{1-\theta}} \ell^{-\gamma} \sum_{j=1}^{J_\ell} j^{1-\gamma}, & \ell = N \\ 0, & N < \ell < \ell_r \\ (1-\delta)K(\gamma-1)T^{\frac{1-\gamma}{1-\theta}} \ell^{-\gamma}, & \ell_r \le \ell \le \ell_{\max} \end{cases}$$

E.2 Estimation

Once again, we adapt the procedure proposed by Garicano et al. (2016) and assume that we only observe labor with measurement error:

$$\ell(\alpha, \varepsilon) = e^{\varepsilon} \ell(\alpha).$$

However, note that $\zeta(\ell)^s$ does not have a closed form integral and has a series of discontinuities between $N/2$ and N. In other words, the distribution of $\ell(\alpha, \varepsilon)$ will not have a closed form solution. Because of this, we will instead numerically compute the firm size density and estimate the underlying parameters using a GMM estimator.

The estimation procedure for this version of the model follows the same steps as the estimation of the firm size distribution without firm splitting: given an initial guess for the parameters, we compute the cumulative distribution of $\ell(\alpha)$, $\mathbb{P}(\ell^*(\alpha) < x)$, by taking the partial sums of $\zeta^s(\cdot)$ and note that

$$\mathbb{P}(\ell(\alpha,\varepsilon) < x|\varepsilon) = \mathbb{P}(\ell^*(\alpha) < e^{-\varepsilon}x|\varepsilon)$$

This conditional distribution can be represented by a matrix

$$\mathbb{P}_{i,j} = \mathbb{P}(\ell < x(i)|\varepsilon(j))$$

where each entry is computed using a two-dimensional grid for x and ε. We then obtain the firm size distribution by integrating $\mathbb{P}_{i,j}$ over $\varepsilon(j)$:

$$\mathbb{P}(\ell < x(i)) = \sum_j \mathbb{P}(\ell < x(i)|\varepsilon(j))\varphi\left(\frac{\varepsilon(j)}{\sigma}\right)\frac{1}{\sigma}.$$

The density ω^s is just the derivative of $\mathbb{P}(\ell < x(i))$,

$$\omega^s(x(i+1)) = \frac{\mathbb{P}(\ell < x(i+1)) - \mathbb{P}(\ell < x(i))}{x(i+1) - x(i)}.$$

Finally, let $s(x)$ be the share of firms who hire x salaried employees in the data. We choose $(\gamma, \sigma, \ell_r, T, \delta)$ to solve

$$\max_{\gamma,\sigma,\ell_r,T,\delta} \quad \sum_{x=1}^{L} [\omega^s(x) - s(x)]^2$$

$$\text{s.t.} \quad (1-\tau_\pi^{\min})\left(\frac{\theta}{1-\theta}\right)T\ell_r \geq \left(\frac{\theta}{1-\theta}\right)\ell_r T^{\frac{1}{1-\theta}}\phi^{\frac{\theta}{1-\theta}} + N\left(\frac{1}{\phi} - (1+\tau_w)\right)$$

where the constraint above is equivalent to equation (8) and $L = 150$. Note that we do not use the percentage deviation between the model and the data in our objective function, which is due to large amount of zeros in our data, especially for larger values of employment.

Due to the complicated nature of ω^s, the solution to the problem above can be sensitive to the initial guess. Thus, we use the ML solution (table 3) as a starting point for $(\gamma, \sigma, \ell_r, T)$ and a number of initial guesses for δ, ranging from 0.01 to 0.25.

E.2.1 Results

Our estimates are given in Table A.3, where the confidence interval is computed by bootstrapping the data 500 times. Note that the estimates do not include $\ell_{\max}$, which is not identified in the GMM version of the model because the numerical density needs to be rescaled to sum up to 1. We therefore use $\ell_{\max} = 23179$, which is the number of salaried employees of the largest firm in the data.

Table A.3: Parameters from GMM estimation

Parameter	Estimated Value	95th CI
γ	1.98	[1.95, 2.01]
σ	0.15	[0.11, 0.20]
ℓ_r	36.0	[35.3, 42.5]
T	1.143	[1.140, 1.147]
δ	0.12	[0.04, 0.13]

Note: Confidence interval computed for a 5% significance level.

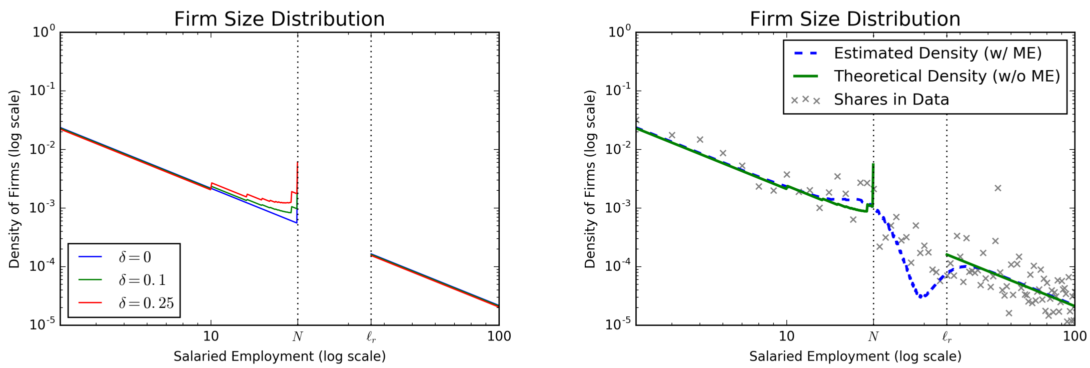
Adding the possibility of firm splitting into our model has two predictions with regards to the firm size distribution. First, the mass of larger firms drops by a proportion of δ. However, as shown in Figure A.2a, this drop is not a substantial one. The main effect of including firm splitting comes from the added mass on the distribution between $N/2$ and N. As evident in Figure A.2a, the distribution still exhibits a mass point at N, but the region just to the right of N has a bulging of the distribution. Furthermore, this bulging is asymmetric, since it does not occur to the right of N. This is a characteristic that we would not obtain from the previous version of the model, and does seem to agree with the data.

It is not surprising that our estimation with firm splitting yields similar results to the estimation without splitting, with two main differences. First, because δ already generates the bulging of the distribution below N, the estimated σ is lower (we no longer need the measurement error to have a large variance to generate bulging). Second, note that both δ and T have the same effect on the density for larger firms, which is to decrease the mass of each of those firms. Since δ and T are "competing" to match this feature in the data, we would expect T to be lower when firms can split. One the other hand, as shown in Figure A.2a, the change generated by δ is not a large one, which explains why our estimated value for T is so close in both versions of the model. The firm size distribution is graphed on Figure A.2b

Figure A.2: Firm size distribution with splitting

(a) Firm size density

(b) Estimated density



Note: Panel (a) shows the firm size density for different values of δ. Panel (b) plots the estimated density when firms can split.

F Wage rigidity

F.1 Estimating ρ

In this section, we show how to estimate the degree of wage rigidity, ρ, introduced in section 5.2. Rearranging Equation (5), the extent of wage rigidity ρ can be expressed as

$$\rho = \frac{w - w_0}{w^* - w_0}$$

Intuitively, ρ can be identified by comparing wage differentials in an economy with and without unemployment. While these wage differentials are not directly observed, the expression above suggests that ρ can be recovered using information on unemployment rates, and then computing the corresponding wage levels using the structural model.

Empirically, the main difficulty lies in recovering the share of the unemployment rate that is attributable to the size-dependent regulations, since unemployment is driven by several other unrelated factors. The solution we adopt is to compare the unemployment rates of Peru and Mexico. We chose Mexico as a benchmark because the size-dependent regulations it has in place do not seem to have a large impact on the firm size distribution (Hsieh and Olken, 2014), and because, belonging to the same region, Mexico is a peer country for Peru for which we can obtain reliable estimates of the structural rate of unemployment from the OECD.

We first compute the average unemployment rate between 2005 and 2015, which is approximately 5.5% in Peru and 4.2% in Mexico. This suggests that the unemployment rate in Peru is 1.3% points higher than it would be without the regulation. We then use the model to compute the wage levels that are consistent with these unemployment rates, and find a value for ρ of about 0.49. That is, the wage adjusts only half of the way down into what would be the full employment wage.

F.2 Calculating counterfactuals under partially rigid wages

In computing the changes displayed in Figure 11, there is a small change in the mechanics of wage rigidity relative to section 5.2. When computing the response to a small variation in the threshold value, say from 20 to 21 employees, we would expect the variation in the wage to be small as well. In fact, for a fixed unemployment rate, the wages in both cases are similar. However, if the unemployment rate in the 20-employee threshold economy is big, the initial wage w_0 can be very different from w^* (i.e. the wage that clears the market with no unemployment in the 21-employee threshold economy). Therefore, the wage predicted by the model can change a great deal after a small change in the primitives of the economy.

To address this, we consider a simple thought experiment: what would happen if we allowed wages to vary one time without any underlying change in the economy? The final wage would then be $w^{**} = w_0 + \rho(w^*_{20} - w_0)$, where w^*_{20} is the wage that clears the market of the 20-employee threshold economy with no unemployment. By looking at this equation, it is clear that $\rho(w^*_{20} - w_0)$ is variation in the wage that happens only because we allowed wages to adjust one more time (with no actual change in the underlying primitives) – which we call "artificial" variation. Our normalization on the mechanics of wage rigidity consists of simply subtracting this artificial variation. The final wage is then described the equation:

$$w = w_0 + \rho(w^* - w_0) - \rho(w^*_{20} - w_0) = w_0 + \rho(w^* - w^*_{20}).$$