Firms, Informality and Development: Theory and evidence from Brazil*

Gabriel Ulyssea†
PUC-Rio

January 14, 2015

Abstract

This paper develops and estimates an equilibrium model where heterogeneous firms can exploit two margins of informality: (i) not register their business, the extensive margin; and (ii) hire workers "off the books", the intensive margin. The model encompasses the main competing frameworks for understanding informality and provides a natural setting to infer their empirical relevance. The counterfactual analysis shows that once the intensive margin is accounted for, aggregate firm and labor informality need not move in the same direction as a result of policy changes. Lower informality can be, but is not necessarily associated to higher GDP, TFP or welfare.

JEL Codes: O17, C54, O12.

*I am indebted to James Heckman, Steven Durlauf and Chang-Tai Hsieh for their guidance and constant encouragement. I would like to thank Ricardo Paes de Barros, Azeem Shaik, Rafael Lopes de Melo, Gary Becker (in memoriam), Ben Moll, Rodrigo Soares, Carlos Henrique Corseuil, Miguel Foguel, Leandro Carvalho e Silvia H. Barcellos for their comments and helpful discussions. Special thanks to Claudio Ferraz, Rafael Dix-Carneiro, Stephane Wolton and Dimitri Szerman for their detailed and extremely helpful comments and feedback. I am also thankful to seminar participants at the Winter Meeting of the Econometric Society, Yale, Johns Hopkins, Chicago, IDB, Einaudi Institute, EESP-FGV, EPGE-FGV, PUC-Rio, IPEA, and the Annual Meeting of the Brazilian Econometric Society for helpful comments and suggestions. Of course, all errors are mine. Financial support from CAPES, IPEA and The University of Chicago is gratefully acknowledged.

†Address: Department of Economics, Pontifícia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente, 225, Gávea. Rio de Janeiro, RJ, 22451-900, Brasil. Email: ulyssea@econ.puc-rio.br.

1 Introduction

The informal sector is a prominent feature of most developing economies.[1] The high levels of informality observed in these countries are likely to have deep economic implications. First, they imply widespread tax avoidance, hindering government's ability to provide public goods. Second, informality may distort firms' decisions along important margins, such as the size of their labor force.[2] Third, it allows less productive (informal) firms to compete with more productive (formal) firms, leading to misallocation of resources and potentially large TFP losses [e.g. Hsieh and Klenow (2009)]. Oppositely, informality can be beneficial to growth as it provides *de facto* flexibility for firms that would be otherwise constrained by burdensome regulations [Meghir et al. (2014)]. Finally, it has been increasingly emphasized that the informal sector might play an important role in shaping welfare consequences from trade liberalization in developing countries.[3] Thus, understanding how the informal sector affects the economy and evaluating the firm-level and aggregate impacts of policies towards informality are central issues in economic development.

To get to these questions, it is crucial to have a clear understanding about the role of informal firms in the economy. There exist three competing views of what this role might be [La Porta and Shleifer (2014)]. The first argues that the informal sector is a reservoir of potentially productive entrepreneurs who are kept out of formality by high regulatory costs, most notably entry regulation. The second view sees informal firms as "parasite firms" that are productive enough to survive in the formal sector but choose to remain informal to earn higher profits from the cost advantages of not complying with taxes and regulations.[4] The third argues that informality is a survival strategy for low skill individuals, who are too unproductive to ever become formal. These views offer very different perspectives on informality and its potential consequences for economic development. Nevertheless, there is no consensus about their empirical relevance and therefore about how important they are for understanding informality [Arias et al. (2010)].

In this paper I propose a new framework that distinguishes two margins of informality: (i) whether firms register and pay entry fees to achieve a formal status, the *extensive*

[1] In Brazil, nearly two thirds of businesses, 40% of GDP and 35% of employees are informal. Similarly, the informal sector accounts for around 50% of the labor force and 41.9% of GDP in Colombia, and 60% of workers and 31.9% of GDP in Mexico (Figure A.1). For information on informal sector's size around the world, see Figure A.1, Schneider (2005) and La Porta and Shleifer (2008, 2014).

[2] For example, if enforcement is tighter for larger firms, there will be incentives to remain small in order to avoid taxes and regulations.

[3] See Goldberg and Pavcnik (2003, 2007), Bosch et al. (2012), Menezes-Filho and Muendler (2011), Dix-Carneiro and Kovak (2014) and Cosar et al. (2014).

[4] The first view dates back to the work of De Soto (1989), while the second view has been put forward by Farrell (2004) and Levy (2008), among others.

margin; and (ii) whether firms that are formal in the first sense hire workers "off the books", the *intensive margin*. The latter is a key innovation, both conceptually and quantitatively. The existing literature has focused on the extensive margin alone, which implies that being informal is a binary decision to comply or not with taxes and regulations.[5] I build on this literature to introduce the intensive margin, which breaks the direct association between firm and worker informality. Accounting for both margins also allows to uncover new and subtler firm-level responses to policy changes regarding informality decisions. I show that these responses translate into non-obvious and quantitatively important effects on TFP, GDP, and aggregate informality. Empirically, I present evidence that the intensive margin accounts for a large share of total informal employment.

In the model, sector membership is defined by the extensive margin, and the (in)formal sector is formed by (un)registered firms. If a firm decides to be formal, it faces fixed entry (registration) costs and higher variable costs due to revenue and labor taxes. However, it may avoid the latter by hiring informal workers. If a firm decides to be informal it avoids all taxes and regulations, but faces an expected cost of being caught that is increasing in firm's size. Since productivity and size are one-to-one in the model, more productive firms (in expectation) self-select into the formal sector and less productive firms enter the informal sector. Having both margins of informality introduces a size-dependent distortion in the economy that is able to rationalize two prominent features of firm size distribution in developing countries: the absence of meaningful discontinuities or bunching of firms at specific points; and the vast predominance of small firms with a small number of medium sized and large firms, even in the formal sector [Hsieh and Olken (2014)]. Additionally, the model predicts some overlap between formal and informal productivity and firm size distributions, as observed in the data.

The proposed model encompasses the three leading views about informal firms discussed above, and is able to integrate them in a unified setting. Even though these views are seen as opposing frameworks, I show that in fact they are not. They simply reflect heterogeneous firms choosing whether to comply given the institutional framework they face. The central distinction lies in their predictions about informal firms' behavior in face of specific policy changes. I exploit these differences to define a taxonomy of informal firms based on these views, which provides a natural setting to infer how important they are in the data.[6] I estimate the model with the simulated method of moments and using three different data sources on formal and informal firms and workers in Brazil. I then

[5] See Rauch (1991), Fortin et al. (1997), Amaral and Quintin (2006), de Paula and Scheinkman (2010, 2011), and Galiani and Weinschelbaum (2012), among others.

[6] The relevant margin to define the taxonomy is the extensive margin. Nevertheless, the intensive margin is central for measuring correctly the relative size of each view in the data and, most importantly, for the effects I find in the counterfactual analysis (discussed ahead).

use the estimated model and the proposed taxonomy to infer the relative size of each view in the data.

The results show that the potentially productive entrepreneurs that are restricted by high bureaucratic entry costs correspond to 16.8% of all informal firms. Those that are productive enough to survive in the formal sector but choose to remain informal to earn higher profits correspond to 38.7% of informal firms. The remaining firms correspond to those too unproductive to ever become formal, which are only able to survive because they avoid taxes and regulations. These results suggest that informal firms are to a large extent "parasite firms" and therefore eradicating them (e.g. through tighter enforcement) could produce positive effects on the economy. Oppositely, given the small fraction of informal firms constrained by entry costs, reducing these would have limited effects on informality and overall economic performance.

In order to assess these conjectures, I use the estimated model to conduct counterfactual analyses of different formalization policies. I consider four prototypical policy interventions: (i) reducing formal sector's entry costs; (ii) reducing the payroll tax; (iii) increasing the cost of the extensive margin of informality through greater enforcement on informal firms (e.g. more government auditing); and (iv) increasing the costs of the intensive margin through tighter enforcement on formal firms that hire informal workers.

At the firm level, the results show that reducing formal sector's entry cost has large positive impacts on informal firms that formalize – an average gain of 24% in terms of their own lifetime profits at baseline – but it has negative effects on other firms. The latter is a consequence of general equilibrium effects: greater entry increases competition and therefore the equilibrium wage increases, hurting incumbents in both sectors. Increasing the costs of the extensive margin of informality benefits formal incumbents, but particularly so low-productivity formal firms. This result thus indicates that these firms are the most directly affected by informal firms' competition. Increasing the costs of the intensive margin of informality is most harmful to low productivity formal firms, as these firms hire a large fraction of their labor force without a formal contract. Thus, they experience a substantial increase in their *de facto* labor cost as a consequence of this policy.

At the aggregate level, reducing formal sector's entry cost leads to a substantial reduction in the share of informal firms but the effect on the share of informal workers is nearly zero. Albeit puzzling at first, this results illustrates the importance of accounting for the intensive margin of informality. Reducing formal sector's entry cost induces low-productivity firms to formalize, which decreases *firm informality*; however, these newly formalized firms hire a large share of informal workers, and therefore the net effect on *labor informality* is nearly null. The opposite is true when increasing enforcement on

the intensive margin: it generates a small reduction in the share of informal workers and actually *increases* informality among firms. The latter effect is observed because the *de facto* cost of being formal increases for less productive firms, as it is now harder for them to hire informal workers, thus increasing their incentives to become informal. These subtler policy impacts can only be uncovered if one explicitly considers the intensive margin. The existing literature has focused on the extensive margin alone, and therefore reducing firm informality necessarily leads to lower labor informality (and vice-versa). As these results show, however, firm and labor informality can move in opposite directions as firms optimally respond to different policies towards informality.

Reducing entry costs also substantially increases the mass of active firms in the economy and leads to greater competition, GDP and wages. Nevertheless, it has a negative effect on aggregate TFP due to negative composition effects, as the mass of active firms increases due to a larger presence of low-productivity firms. Increasing enforcement on the extensive margin nearly eradicates informal firms, which generates a large positive effect on aggregate TFP. Even though higher aggregate TFP goes in the direction of increasing production, the substantial reduction in the mass of active firms goes in the opposite direction, and GDP remains roughly unchanged. This policy also generates a positive but small effect on welfare, which is entirely driven by a substantial increase in tax revenues. Overall, the welfare analysis shows that even though the policies analyzed always reduce at least one margin of informality, they do not necessarily lead to welfare improvements.

The firm-level results are related to a literature stream that uses micro data to analyze the impact of different formalization policies in developing countries, among others: Monteiro and Assunção (2012) and Fajnzylber et al. (2011), who analyze tax reduction and simplification; Bruhn (2011), Kaplan et al. (2011) and De Mel et al. (2013), who analyze the effects of reducing formal sector's bureaucratic entry costs; Rocha et al. (2014), who separately estimate the impacts of reducing entry costs and taxes; Almeida and Carneiro (2009, 2012) and de Andrade et al. (2013), who analyze the impacts of greater government auditing. The present approach, however, allows me to compute the full distribution of firm-level effects and to account for general equilibrium effects, which I show to be sizable. This paper is also related to the literature that analyzes aggregate effects of policies towards informality, which include Ulyssea (2010), Prado (2011), Charlot et al. (2011), D'Erasmo and Boedo (2012), and Leal Ordonez (2014), among others. The present framework embeds firm behavior into aggregate relationships, and thus allows to simultaneously assess policy impacts on firm-level and aggregate outcomes, which have been separately analyzed by these literature streams. A notable exception is the recent work by Meghir et al. (2014), who develop a wage-posting model with formal and

informal sectors. Search frictions play a central role in their analysis, which is based on individual worker data from Brazil. Their focus lies on the analysis of labor markets and their approach can thus be seen as complementary to the one proposed in this paper.

The remaining of the paper is organized as follows. Section 2 presents the data and some key stylized facts. Section 3 presents the model, while Section 4 discusses the taxonomy of informal firms. Section 5 contains the estimation method and results. Section 6 presents the quantitative results and Section 7 concludes.

2 Facts about firm informality

2.1 Definitions and Data

Throughout this paper, I define as *informal workers* those employees who do not hold a formal labor contract, which in Brazil is defined by having a booklet (*carteira de trabalho*) that registers workers' entire employment history in the formal sector. I define as *informal firms* those not registered with the tax authorities, which means that they do not possess the tax identification number required for Brazilian firms (*Cadastro Nacional de Pessoa Juridica* – CNPJ). These definitions are used in the theory as well as in the data.

I use four data sets to conduct the empirical analysis. The two main ones are those that contain information of formal and informal firms in Brazil. The first is the ECINF survey (*Pesquisa de Economia Informal Urbana*), a repeated cross-section of small firms (up to five employees), which was collected by the Brazilian Bureau of Statistics (IBGE) in 1997 and 2003. This is a matched employer-employee data set that contains information on entrepreneurs, their business and employees. Firms are directly asked whether they are registered with the tax authorities and whether each of their workers has a formal labor contract. Thus, it is possible to directly observe firms' status as well as their workers'.[7] The ECINF is designed to be representative at the national level for firms with at most five employees.[8]

Although ECINF's sample size cap is not likely to be a problem when analyzing in-

[7]These are self-reported variables and naturally raise measurement error concerns. Nonetheless, the National Bureau of Statistics (IBGE) has a long tradition in measuring labor informality with high accuracy, and it has very strict confidentiality clauses, so the information cannot be used for auditing purposes (which could incentivize respondents to misreport). These features, associated to the actual high levels of informality observed in the data, increase the confidence that respondents are not deliberately underreporting their informality status.

[8]The effective sample includes firms with up to 10 employees, but the information for larger firms is not representative at a national level. See de Paula and Scheinkman (2010) for a more detailed description of the ECINF data set.

formal firms, which are predominantly small scale enterprises, it certainly is a binding restriction for the analysis of formal firms. I therefore use the RAIS data set to complement the information on formal firms. This is an administrative data set collected by the Ministry of Labor, which provides an annual panel with the universe of formal firms and workers. Having these two data sets also allows me to assess the quality of the data in the ECINF, comparing it with the administrative records in RAIS. As Table 1 shows, both the size distribution and the composition across industries is remarkably similar in RAIS (restricted to firms with up to 5 employees) and ECINF, which is reassuring of ECINF's quality.[9]

Finally, I also use two household surveys collected by the Brazilian Bureau of Statistics to compute some aggregate labor market statistics (such as the share of informal workers). The first is the National Household survey (PNAD), a repeated cross section that is representative at the national level. The second is the Monthly Employment Survey (PME), which is a rotating panel of workers that covers the 6 main metropolitan areas in Brazil.

Table 1: Comparing ECINF and RAIS

	RAIS (size ≤ 5)	Formal – ECINF	Informal – ECINF
Sector composition (%)			
Services	40.9	42.5	53.7
Manufacturing	9.6	7.9	8.9
Commerce	47.2	49.6	37.4
Size Distribution (# workers)			
Pc. 25	1	1	1
Pc. 50	2	2	1
Pc. 75	3	3	1
Pc. 95	5	5	3
Mean	2.2	2.1	1.3
Obs.	1,570,105	2,600	18,736

Source: Author's own tabulations from RAIS and ECINF, 2003.

2.2 Facts

There exist some well-established facts about informal firms in the literature [e.g. Perry et al. (2007) and La Porta and Shleifer (2008)]: on average they have less educated entrepreneurs, are smaller both in terms of employees and revenues, pay lower wages and

[9] Appendix B describes the details of the construction of the data sets used.

earn lower profits relatively to formal firms. These facts are also present in the Brazilian data [e.g. de Paula and Scheinkman (2011)]. These stark differences between formal and informal firms have been often interpreted as evidence that they operate in completely separate industries and produce entirely different products. However, Figure B.1 in the appendix provides evidence that they coexist even within narrowly defined industries (at the 7-digit level), which contradicts the notion that formal and informal firms operate in completely different markets.

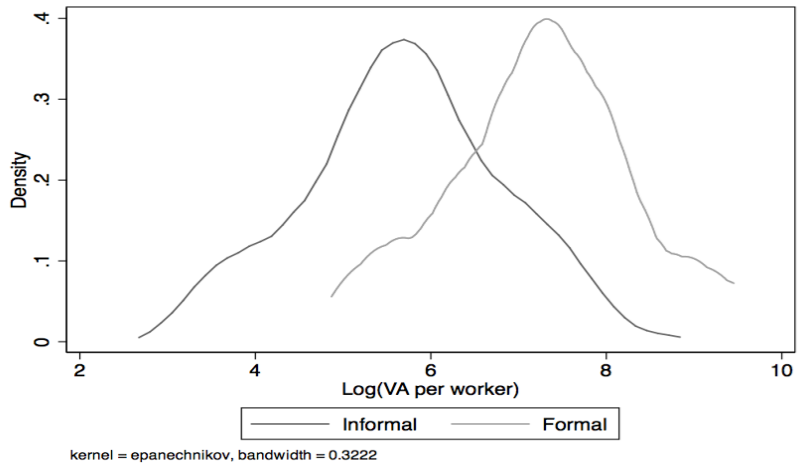
It has also been shown that these observed differences in average outcomes between formal and informal firms reflect substantial differences in average productivity. I take a step further and ask to what extent these differences are due to firms sorting into both sectors based on productivity right upon entry. For that, I compute proxies for productivity (value-added per worker) and size (log-revenues) for formal and informal firms at most one year old to proxy for entrants.[10] Figure 1 shows that both productivity and size distributions in the formal sector are already substantially shifted to the right among very young firms, which is consistent with firms sorting based on productivity right upon entry. Moreover, there is a large overlapping region in both firm size and productivity distributions, which is a largely overlooked empirical regularity. Thus, not only formal and informal firms produce in the same industry but there is also a sizable interval in the productivity support where one can find both types of firms. Meghir et al. (2014) show that this overlapping region is also present if one considers all formal and informal firms (and not only the entrants).

As for the two margins of informality, a well-known fact in the literature is that the probability of being informal (the extensive margin) strongly decreases with firms' size, usually measured as number of employees [e.g. Perry et al. (2007)]. The same pattern is observed in the Brazilian data [see Figure B.2 in the appendix and de Paula and Scheinkman (2011)]. One possible rationale behind this fact is simply that larger firms are too visible to the government and thus more likely to be audited. Given this argument, it is likely that the same pattern would be observed for the intensive margin: larger formal firms (in number of employees) should have a lower share of informal employees. Indeed, I provide evidence that the intensive margin of informality is decreasing in firm's size (Figure B.2 in the appendix).

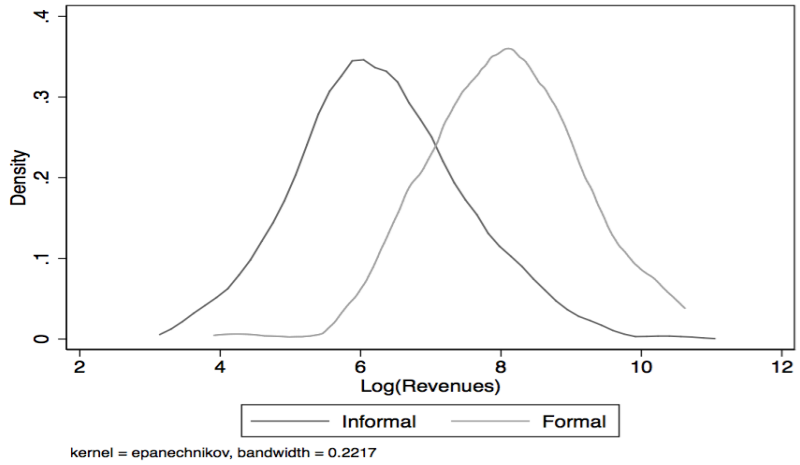
Finally, I assess the empirical relevance of the intensive margin, which can only be done indirectly with the data available. In Table 2, I use data from the Monthly Employment Survey (PME) to show that 52% of all informal workers are employed in firms with

[10] To obtain cleaner measures, I regress the log of value-added and log-revenues on a set of industry dummies to purge inter-industry variation. The computed log-residuals are the productivity and size measures used.

Figure 1: Productivity and size distributions among entrants



(a) Productivity: Log(VA/Worker)



(b) Size: Log(Revenues)

Notes: Data from ECINF. I regress the log of value-added per worker and log-revenues on a set of industry dummies to purge inter-industry variation. The figures show the densities of computed log-residuals for formal and informal firms.

11 employees or more. However, as already discussed, the likelihood of a firm with 11 employees or more to be informal is very low. These two pieces of evidence combined thus suggest that there is a large fraction of informal workers who are employed in formal firms.

Table 2: Formal and informal employment composition by firm size

	Informal Workers (in %)	Formal workers (in %)
Firm size (# employees)		
0–5	35.8	6.6
6–10	11.7	7.2
11 or more	52.5	86.2

Source: Author's own tabulations from the Monthly Employment Survey (PME) 2003.

3 Theory

Motivated by the facts previously discussed, this section develops an equilibrium entry model where firms can exploit both the extensive and intensive margins of informality. Firms are heterogeneous and indexed by their individual productivity, θ. Firms produce a homogeneous good using labor as their only input. Product and labor markets are competitive, and formal and informal firms face the same prices.[11] To simplify the exposition, I assume that workers are homogeneous and formal and informal employees perform the exact same task within the firm.[12]

3.1 Incumbents

Incumbents in both sectors have access to the same technology. Output of a given firm θ is given by $y(\theta, \ell) = \theta q(\ell)$, where the function $q(\cdot)$ is a assumed to be increasing, concave, and twice continuously differentiable.

Informal incumbents are able to avoid taxes and labor costs, but face a probability of detection by government officials. This expected cost takes the form of a labor distortion denoted by $\tau_i(\ell)$, where $1 \leq \tau_i(\cdot) < \infty$, and it is assumed to be increasing and convex in firm's size ($\tau_i', \tau_i'' > 0$). These assumptions can be rationalized, for instance, by the fact that larger firms have a greater probability of being caught [e.g. Fortin et al. (1997), de Paula and Scheinkman (2011) and Leal Ordonez (2014)].[13] Informal firms' profit

[11] As argued in Section 2, formal and informal firms coexist even within narrowly defined industries, so the assumption that firms face the same output price seems like a reasonable approximation. Nevertheless, the model can be readily modified to a monopolistic competition setting where firms produce different varieties.

[12] Worker homogeneity is a strong assumption, which implies among other things that formal and informal workers receive the same market wage. Nevertheless, as long as there is positive assortative matching between workers and firms, the results here remain largely unaltered.

[13] The general cost function $\tau_i(\cdot)$ can be directly obtained from a formulation that explicitly accounts for a detection probability [see Ulyssea (2014)].

function is thus given by:

$$\Pi_i(\theta, w) = \max_{\ell} \{\theta q(\ell) - w\tau_i(\ell)\} \tag{1}$$

where the price of the final good is normalized to one.

Formal incumbents must comply with taxes and regulations, but they can hire informal workers to avoid the costs implied by the labor legislation. The hiring costs of formal and informal workers differ due to institutional reasons: formal firms have to pay a constant payroll tax on formal workers, while they face an increasing and convex expected cost to hire informal workers, which is summarized by the function $\tau_{fi}(\cdot)$, $\tau'_{fi}, \tau''_{fi} > 0$ and $1 \leq \tau_{fi}(\cdot) < \infty$. The cost for formal firms of hiring informal workers is thus given by $\tau_{fi}(\ell)w$, while the cost of hiring formally is $(1+\tau_w)w$, where τ_w is the labor tax. Since formal and informal workers are perfect substitutes, on the margin firms hire the cheapest one, and hence there is an unique threshold $\tilde{\ell}$ above which formal firms only hire formal workers (on the margin).[14] Formal firms' profit function can be written as follows:

$$\Pi_f(\theta, w) = \max_{\ell} \{(1-\tau_y)\theta q(\ell) - C(\ell)\} \tag{2}$$

and

$$C(\ell) = \begin{cases} \tau_{fi}(\ell)w, & \text{for } \ell \leq \tilde{\ell} \\ \tau_{fi}\left(\tilde{\ell}\right)w + (1+\tau_w)w\left(\ell - \tilde{\ell}\right), & \text{for } \ell > \tilde{\ell} \end{cases} \tag{3}$$

where τ_y denotes the revenue tax. Incumbents in both sectors must pay a per-period, fixed cost of operation, which is denoted by $\bar{c}_s$, $s = i, f$. This a standard formulation in the literature and can be interpreted as the opportunity cost of operating in sector s. The profit function net of this fixed cost of operation is denoted by $\pi_s(\theta, w) = \Pi_s(\theta, w) - \bar{c}_s$.

The two margins of informality introduce a size-dependent distortion in the economy, as lower productivity (smaller) firms face *de facto* lower marginal costs. By the same argument, more productive, larger firms are more likely to be formal, as the costs of the extensive margin of informality are increasing in firm's size ($\tau'_i, \tau''_i > 0$). Since formal firms only hire formal workers in excess of $\tilde{\ell}$, the share of informal workers within a formal firm is also monotonically decreasing in firm's size (as observed in the data). Thus, this

[14] The marginal cost of hiring informal workers $\left[w\tau'_{fi}(\ell)\right]$ is strictly increasing, while the marginal cost of hiring formal workers $[(1+\tau_w)w]$ is constant. Hence, there is an unique value of ℓ such that $\tau'_{fi}(\tilde{\ell}) = 1+\tau_w$. If the labor quantity that maximizes formal firm's profit is such that $\ell^* \leq \tilde{\ell}$, then it will only hire informal workers. If $\ell^* > \tilde{\ell}$, then the firm will hire $\tilde{\ell}$ informal workers and $\ell^* - \tilde{\ell}$ formal workers.

highly tractable formulation is able to capture the main facts discussed in the previous section regarding both margins of informality.[15]

3.2 Entry

Every period there is a large mass of potential entrants of size M. Potential entrants only observe a pre-entry productivity parameter, $\nu \sim G$, which can be interpreted as a noisy signal of their effective productivity. G is assumed to be absolutely continuous with support $(0, \infty)$, with finite moments, and it is the same for all firms and independent across periods (i.e. ν is i.i.d.). Hence, the mass of entrants in one period does not affect the composition of potential entrants in the following period. To enter either sector, firms must pay a fixed cost (denominated in units of output) that is assumed to be higher in the formal sector: $E_f > E_i$.[16] After entry occurs, firms draw their actual productivity from the conditional c.d.f. $F(\theta|\nu)$, which is the same in both sectors and independent across firms. $F(\theta|\nu)$ is assumed to be continuous in θ and ν, and strictly decreasing in ν. Hence, a higher ν implies a higher probability of a good productivity draw after entry occurs.

Since firms are *ex ante* heterogeneous but only realize their actual productivity after entry occurs, the model allows for the possibility of overlap between formal and informal productivity distributions. Oppositely, the fully static models without uncertainty imply perfect sorting and no overlap between formal and informal firms' productivity and size distributions,[17] which is at odds with the data (as shown in Section 2).

If firms are surprised with a low productivity draw $\theta < \bar{\theta}$, where $\pi_s\left(\bar{\theta}, w\right) = 0$, they decide to exit immediately without producing. If firms decide to stay, their productivity remains constant forever and they face an exogenous exit probability denoted by κ_s, $s = i, f$. Note that this exit probability could also be interpreted as a sector-specific discount rate, which could reflect, for example, differential borrowing rates. Aggregate prices remain constant in steady state equilibria and since firms' productivity also remains

[15] However, this formulation also implies that all formal firms hire some informal workers (up to $\tilde{\ell}$), which for very large firms might be unrealistic (even though their share of informal workers is going to zero). Nevertheless, given that the model captures well the behavior of the *share* of informal workers within formal firms, the gains in tractability justify this modeling option.

[16] The difference between these entry costs is interpreted here as a consequence of the regulation of entry into the formal sector (e.g. red tape bureaucracy). Under this interpretation, the entry cost into the informal sector can be seen as the initial investment or minimum scale required to operate in the given industry.

[17] See, for example, Rauch (1991), Fortin et al. (1997), de Paula and Scheinkman (2011), Prado (2011), and Galiani and Weinschelbaum (2012).

constant, firm's value function assumes a very simple form:

$$V_s(\theta, w) = \max\left\{0, \frac{\pi_s(\theta, w)}{\kappa_s}\right\}$$

where for notational simplicity I assume that the discount rate is normalized to one. The expected value of entry for a firm with pre-entry signal ν is thus given by

$$V_s^e(\nu, w) = \int V_s(\theta, w)\, dF(\theta|\nu), \qquad s = i, f \tag{4}$$

Entry into the formal sector occurs if $V_f^e(\nu, w) - E_f \geq \max\{V_i^e(\nu, w) - E_i, 0\}$, while entry into the informal sector occurs if $V_i^e(\nu, w) - E_i > \max\left\{V_f^e(\nu, w) - E_f, 0\right\}$. If entry in both sectors is positive the following entry-conditions hold:

$$\begin{aligned} V_i^e(\overline{\nu}_i, w) &= E_i \\ V_f^e(\overline{\nu}_f, w) &= V_i^e(\overline{\nu}_f, w) + (E_f - E_i) \end{aligned}$$

where $\overline{\nu}_s$ is the pre-entry productivity of the last firm to enter sector $s = i, f$. The appendix D.1 shows that the effective, post-entry productivity distributions in both sectors can be derived as functions of these thresholds.

3.3 Equilibrium

To close the model, it is necessary to specify the demand side of the model. I assume that there is a representative household that inelastically supplies $\overline{L}$ units of labor and that derives utility solely from consuming the final good, x:

$$U = \sum_{t=0}^{\infty} \beta^t u(x_t)$$

The focus lies on stationary equilibria, where all aggregate variables remain constant. In a stationary equilibrium with constant prices the household problem simplifies to a static optimization problem:

$$\max_x u(x) \qquad \text{s.t.} \quad x \leq w\overline{L} + \Pi + T \tag{5}$$

The Π denotes total profits in the economy net of total entry costs, $M_f E_f + M_i E_i$, where $M_i = [G(\overline{\nu}_f) - G(\overline{\nu}_i)]\, M$ and $M_f = [1 - G(\overline{\nu}_f)]\, M$ denote the measures of entrants into the informal and formal sectors, respectively. The T denotes tax revenues, which are directly transferred to households. Consumers do not derive any disutility from

work and cannot save, so they simply consume all their income. Total consumption thus constitutes the natural welfare measure in this context, which in equilibrium is given by $W = w\overline{L} + \Pi + T$.

In a stationary equilibrium, the size of the formal and informal sectors must remain constant over time, which implies the following condition:

$$\mu_s = \frac{1 - F_{\theta_s}\left(\overline{\theta}_s\right)}{\kappa_s} M_s \tag{6}$$

where μ_s denotes the mass of active firms in sector s. In words, condition (6) simply states that the mass of successful entrants in both sectors must be equal to the mass of incumbents that exit.

In sum, the equilibrium conditions are given by the following: (i) markets clear, $L_i + L_f = \overline{L}$; (ii) The zero profit cutoff (ZPC) condition holds in both sectors, $\theta \geq \overline{\theta}_s$ where $\pi_s\left(\overline{\theta}_s, w\right) = 0$; (iii) the free entry condition holds in both sectors, with equality if $M_s > 0$; and (iv) both sectors' size remains constant (expression 6). Appendix D.2 shows that the equilibrium exists and it is unique.

4 A Taxonomy of Informal Firms

In this section I propose a simple taxonomy of informal firms based on the three main competing views that exist in the literature [see La Porta and Shleifer (2008, 2014) for a discussion]. The starting point of the analysis is to establish a precise definition of each type, which comes directly from these views:

- Survival view (type 1): Informal firms that are too unproductive to ever become formal, even if entry costs were removed. These are entrepreneurs with low human capital, who are only able to survive in the informal sector because they avoid taxes and regulations.

- Parasite view (type 2): Informal firms that are productive enough to survive as formal firms once entry barriers are removed, but *choose* not to do so because it is more profitable for them to remain informal.

- De Soto's view (type 3): Higher productivity informal firms that are kept out of formality by high entry costs. If these were removed, they would become formal and improve their performance, as they would no longer have the size constraints imposed by informality.

The crucial difference between these types is how they would respond to a policy that eliminates entry costs into the formal sector. Type 3 firms would formalize their business and would be better off in this counter-factual scenario, as they are no longer constrained by the growth limitations imposed by informality. Hence, any model that does not account for entry costs into the formal sector cannot account for this view, as there would be no bunching of informal firms near the transition threshold. However, the other two types are not so easily distinguishable, as both are predicted to remain informal in the absence of entry costs. Any model that has firms sorting between sectors, even without entry costs and productivity uncertainty, would be able to account for these two types. The crucial differentiation between Types 1 and 2 is the reason why they remain informal. Type 2 firms are productive enough to survive in the formal sector (once entry barriers are removed), but choose to remain informal to receive higher profits. Type 1 firms are simply not productive enough to be formal, and are only able to survive due to the cost advantages of non-compliance.

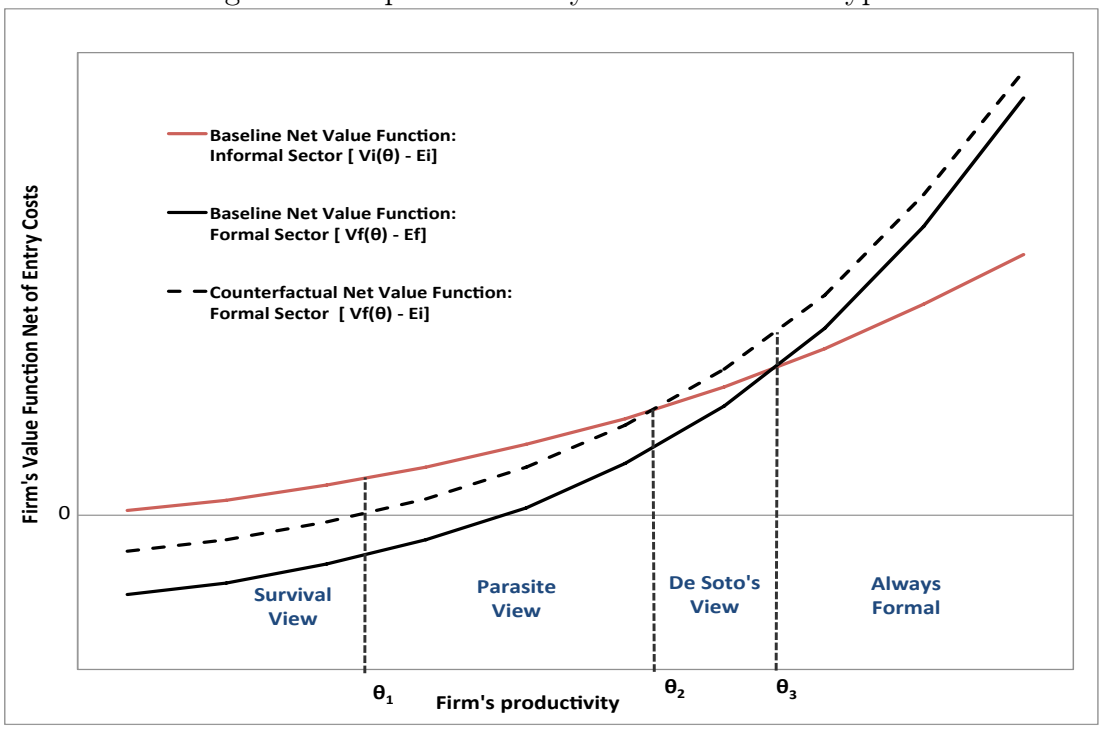
Given this reasoning, it is possible to represent the different types using the model just discussed. The thought experiment is to ask how informal firms would respond to an intervention that equalizes entry costs in the formal and informal sectors ($E_f = E_i$). To disentangle types 1 and 2 it is necessary to ask a somewhat harder question: Which firms could actually become formal in the counter-factual scenario but choose not to do it? Figure 2 summarizes this thought experiment. For each post-entry productivity level (θ),[18] it plots firm's post-entry value function net of entry costs in both sectors at baseline, $V_s(\theta) - E_s$, and the value of being formal under the counter-factual scenario where formal and informal sectors' entry costs are equalized, $V_f^c(\theta) - E_i$, where the superscript c indicates the counter-factual scenario.[19]

The baseline curves for the formal and informal sectors intersect each other at $\theta = \theta_3$, and all firms with $\theta \geq \theta_3$ will always choose to be formal, as their value is higher than being informal. Firms with productivity $\theta \in [\theta_2, \theta_3)$ are the De Soto's (type 3) firms: once entry barriers into the formal sector are removed, they migrate to the formal sector, improve their performance and achieve higher profits. Firms with productivity $\theta \in [\theta_1, \theta_2)$ correspond to Parasite (type 2) firms: They are productive enough to produce in the formal sector – their value in the formal sector is everywhere above zero – but choose not to do it to obtain higher returns in the informal sector. Finally, firms with $\theta < \theta_1$ are the Survival (type 1) firms, which are not productive enough to go to the formal sector even

[18] Firms make entry decisions and sort between sectors based on their pre-entry signal, ν, and after entry occurs they realize their actual productivity, θ, which is drawn from $F(\theta|\nu)$.

[19] To obtain these curves it is necessary to specialize the model to specific functional forms and parameter values. I postpone this discussion to the following section, where I present the estimation procedure.

Figure 2: Graphic taxonomy of informal firms types



Note: The figure shows, for each productivity level, firms' post-entry baseline value function net of entry costs in the formal and informal sectors, $V_f(\theta) - E_f$ and $(V_i(\theta) - E_i)$, respectively. The third curve displays the post-entry, net value of being formal in the counter-factual scenario where formal sector's entry costs are equated to informal sector's $(E_f = E_i)$: $V_f^c(\theta) - E_i$, where the superscript c indicates the counter-factual scenario.

without entry costs, and use informality as a survival strategy.

Contrary to what is often argued in the literature,[20] Figure 2 shows that these views are not fundamentally different, they simply reflect firm heterogeneity. Thus, they are complementary and not competing frameworks for understanding informality. The crucial question is therefore to infer the relative importance of each view in the data. For that, it is necessary to estimate the model and use it to back out the mass of firms in each of the θ intervals just described. The following section describes the estimation procedure, while Section 6 describes the quantitative results.

5 Estimation

The model presented in Section 3 describes firms' decisions regarding entry, production and compliance with regulations in an equilibrium setting. To perform counter-factual analysis and make quantitative statements about firms types and how they would respond

[20] See La Porta and Shleifer (2008, 2014) for a systematic overview of the existing literature regarding these views.

to policy changes, it is necessary to estimate all objects in the model's structure. I estimate the model using a two-stage simulated method of moments (SMM). This approach combines direct estimation and calibration from micro and macro data in the first stage, with the SMM estimator itself in the second stage [e.g. Gourinchas and Parker (2002)].

To proceed with the estimation, it is first necessary to complete the model's parameterization and assume functional forms for the different objects in the model,[21] which naturally raises identification concerns. However, non-parametric estimation in the present context is either not feasible (given the goals of this paper and the data available), or the assumptions needed are not attainable.[22] The fully parametric approach adopted here is thus crucial in order to overcome some data limitations and to provide a rich counter-factual analysis.

5.1 Parameterization

Up to this point, the initial productivity distribution, G_ν, the productivity process, $F(\theta|\nu)$, the production function, $q(\cdot)$, and the cost functions, $\tau_s(\cdot)$, were left unspecified. This section completes the model's parameterization by assuming specific functional forms for these objects. Starting with the pre-entry productivity distribution, it is assumed that it has a Pareto distribution:[23]

$$F_\nu\left(\nu \geq x\right) = \begin{cases} \left(\frac{\nu_0}{x}\right)^\xi & \text{for } x \geq \nu_0 \\ 1 & \text{for } x < \nu_0 \end{cases} \tag{7}$$

Firms' actual productivity is only determined after entry occurs. I assume a very simple log-additive form for the post-entry productivity process, which is determined as follows: $\theta = \varepsilon\nu$, where the unexpected shock ϵ is i.i.d. and has a log-normal distribution with mean zero and variance σ^2. As for production, I assume that firms use a Cobb-Douglas technology: $y\left(\theta, l\right) = \theta l^\alpha$, $\alpha < 1$.

The cost functions of both margins take a very simple functional form: $\tau_s(\ell) = \left(1 + \frac{\ell}{b_s}\right)\ell$, where $b_s > 0$ and $s = i, f$. Finally, I assume that the per-period, fixed costs of operation are a function of the equilibrium wage, which makes the exit margin more meaningful since it now responds to market conditions. The fixed costs are

[21] It is not always the case that one needs to identify all the objects in the model's structure in order to answer specific policy questions [see, for example, Heckman (2001) and Ichimura and Taber (2002)]. However, *ex ante* policy evaluations typically require the full specification of a behavioral model in order to perform the counter-factual analysis [e.g. Keane et al. (2011)].

[22] See the Web Appendix for a detailed discussion.

[23] A well documented fact in the literature is that the Pareto distribution fits firms' size distribution remarkably well, see more recently Luttmer (2007), among others.

determined as follows: $\bar{c}_s = \gamma_s w$, $0 < \gamma_s \leq 1$.

The parameter vector to be estimated is thus given by

$$\Gamma = \{\tau_w, \tau_y, \kappa_f, \kappa_i, \alpha, \sigma, \gamma_f, \gamma_i, b_i, b_f, \xi, \nu_0, E_f, E_i\}$$

which is partitioned into two sub-vectors, $\Gamma = \{\psi, \varphi\}$, the first and second stage parameters, respectively. The SMM estimator used in the second stage takes the parameters in ψ as given in order estimate the vector φ. The next subsection describes the estimation steps.

5.2 Fitting the model to the data

The vector of parameters determined in the first stage is given by

$$\psi = \{\tau_w, \tau_y, \kappa_f, \nu_0, \gamma_f\}$$

The tax rates are set to their statutory values: $\tau_w = 0.375$ and $\tau_y = 0.293$.[24] The exit probability in the formal sector is $\kappa_f = 0.129$, which is estimated using the panel structure in the RAIS data set. This estimate is obtained using the predicted exit probability for the average firm in the sample. The Pareto distribution scale parameter (ν_0) is set so that the firms' minimum size is one employee, while formal sector's fixed cost of operation ($\gamma_f = 0.5$) is set to be half of the monthly wage.

5.2.1 Second stage: SMM estimation

For a given parameter vector (Γ), wage (w) and individual productivity shocks (ν_j and ε_j), one can use the model to completely characterize firms' behavior. The SMM estimator proceeds by using the model to generate simulated data sets of formal and informal firms and computing a set of moments that are also computed from real data. The estimate is obtained as the parameter vector that best approximates the moments computed from the simulated data to the ones computed from real data.[25]

Let $\hat{m}$ denote the vector of moments computed from data, and let $m^s(\varphi; \psi)$ denote the vector of the same moments computed from the simulated data. Define $g(\varphi; \psi) = \hat{m} - m^s(\varphi; \psi)$; the SMM estimation is based on the moment condition $E[g(\varphi_0; \psi_0)] = 0$,

[24] The value of τ_w corresponds to the main payroll taxes, namely, employer's social security contribution (20%), direct payroll tax (9%), and severance contributions (FGTS), 8.5%. The value of τ_y includes two VAT-like taxes: the IPI (20%) and PIS/COFINS (9.25%). These values can be easily obtained in the compilation by the World Bank's Doing Business initiative.

[25] The technical details are discussed in the Appendix E.1. The interested reader can find an in-depth and systematized discussion in Gourieroux and Monfort (1996) and Adda and Cooper (2003).

where φ_0 and ψ_0 denote the true values of φ and ψ, respectively. The second-stage, SMM estimator is then given by

$$\hat{\varphi} \quad = \quad \arg\min_{\varphi} Q(\varphi;\psi) = \left\{ g(\varphi;\psi)' \hat{\mathbf{W}} g(\varphi;\psi) \right\} \tag{8}$$

where $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$, and $\mathbf{W}$ is a positive semi-definite weighting matrix. Under the suitable regularity conditions (which are discussed in the web appendix), the SMM estimator is consistent and asymptotically normal. The Appendix E.1 describes the derivation of the asymptotic variance-covariance matrix and the computation of the optimal $\hat{\mathbf{W}}$, while the appendix E.2 describes the optimization algorithm.

5.2.2 Moments and identification

There are 9 (nine) parameters to be estimated in the second stage:

$$\varphi = \{\kappa_i, \gamma_i, b_i, b_f, \xi, \alpha, E_f, E_i, \sigma\}$$

I use 18 moments from the data to form the vector $\hat{m}$,[26] which are the following: (i) share of informal employees (data source: PNAD); (ii) overall share of informal firms and by firm size for $n = 1, ..., 5$, where n is the number of employees (data sources: ECINF and RAIS); (iii) average share of informal workers within formal firms with size $n = 2, ..., 5$ (data source: ECINF); (iv) the 75^{th}, 95^{th} and 99^{th} percentiles of informal firms' size distribution (for firms with up to 5 employees – data from ECINF); and (v) the 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of formal firms' size distribution (data source: RAIS).

Even though simulation-based methods typically do not allow for formal identification arguments, it is possible to discuss the role that some moments play in the identification of different parameters [e.g. Dix-Carneiro (2014)]. The shape parameter of the Pareto distribution, ξ, is completely determined by the moments of firm size distribution. In fact, given the one-to-one relationship between productivity and firms' size in the model, estimation of the productivity distribution crucially relies on observed moments of firm size distribution (measured by number of employees).

As for the parameters that govern the cost functions of both margins of informality,

[26] As discussed in Section 2, the data sets used are the following: ECINF (*Pesquisa de Economia Informal Urbana*), a repeated cross-section of small firms (up to five employees), collected by the Brazilian Bureau of Statistics (IBGE); RAIS (*Relação Anual de Informações Sociais*), an administrative data set from the Ministry of Labor, which provides an annual panel with the universe of formal firms and workers; and PNAD, the National Household Survey, a repeated cross section that is representative at the national level.

the share of informal firms by firm size plays a crucial role in identifying b_i, while the share of informal workers in formal firms by firm size identifies b_f. As for informal sector's exit (or discount) rate κ_i, it determines the overall disadvantage of being informal relatively to being formal, because a higher κ_i represents an overall downward shift in informal firms' value function. Thus, given formal sector's entry cost, κ_i is disciplined by the overall share of informal firms. Given γ_f (determined in the first stage), γ_i and the post-entry shock (σ) are largely determined by the degree of overlap between formal and informal firm size distributions, and the minimum size of entrants in the informal sector. Formal sector's entry cost (E_f) is disciplined by the minimum entry size among formal firms and how shifted to the right firm size distribution is relatively to informal sector's. Similarly, informal sector's entry cost is likely to be determined by entrants' minimum scale, which comes from the lower percentiles of firm size distribution in the informal sector.

5.2.3 Estimates and Model Fit

Table 3 shows the values of both first and second stage parameters, ψ and φ respectively. The estimates show that formal sector's entry cost is more than twice informal sector's. Exit (discount) rate in the informal sector is also more than twice as high as formal sector's, which confirms the anecdotal evidence that informal firms have higher turnover rates than their formal counterparts. Pareto's shape parameter, ξ, indicates that the productivity distribution is skewed to the left, as observed in the data for firm size.

As for the fit of the model, Table 4 shows how the model performs compared to the observed moments in the data. The model matches the share of informal firms and the share of informal workers well. However, it understates the share of informal firms with only one employee (the 75% percentile of the actual size distribution in the informal sector). The same does not happen with the size distribution in the formal sector, which the model is able to replicate well. The main reason for this greater accuracy in estimating formal sector's firm size distribution is that the corresponding empirical moments are more precisely estimated using RAIS, and therefore they receive larger weights in the optimal weighting matrix $\hat{\mathbf{W}}$ (see the discussion in Appendix E.1).

Table 3: Parameter Values

Parameter	Description	Source	Value	SE
First Stage				
τ_w	Payroll Tax	Statutory values	0.375	–
τ_y	Revenue Tax	Statutory values	0.293	–
κ_f	Formal Sector's Exit Probability	Panel Estimation	0.129	–
ν_0	Pareto's Location Parameter	Calibrated	7.7	–
γ_f	Per-period fixed cost of operation (Formal)	Calibrated	0.5	–
Second Stage				
α	Cobb-Douglas Coefficient	Estimated (SMM)	0.649	0.003
b_f	Intensive Mg. Cost: $\tau_{fi}(\ell) = \left(1 + \frac{\ell}{b_f}\right) \ell$	Estimated (SMM)	4.592	0.127
b_i	Extensive Mg Cost: $\tau_i(\ell) = \left(1 + \frac{\ell}{b_i}\right) \ell$	Estimated (SMM)	4.522	0.038
κ_i	Informal Sector's Exit Probability	Estimated (SMM)	0.349	0.017
γ_i	Per-period fixed cost of operation (Informal)	Estimated (SMM)	0.246	0.035
ξ	Pareto's Shape Parameter	Estimated (SMM)	3.9	0.064
σ	Post-Entry Shock Variance	Estimated (SMM)	0.141	0.005
E_f †	Formal Sector's Entry Cost	Estimated (SMM)	6,077	372
E_i †	Informal Sector's Entry Cost	Estimated (SMM)	2,550	177.2

† Estimates and SD expressed in R$ of 2003.

Table 4: Model fit

Moments	Source	Data	Model
Share inf. workers	PNAD	0.354	0.352
Share inf. Firms	ECINF & RAIS	0.686	0.695
Size Distribution: Informal Firms			
≤ 1 *employee*	ECINF	0.849	0.481
≤ 2 *employees*	ECINF	0.958	0.946
≤ 4 *employees*	ECINF	0.993	0.995
Size Distribution: Formal Firms			
≤ 1 *employee*	RAIS	0.295	0.299
≤ 3 *employees*	RAIS	0.563	0.545
≤ 7 *employees*	RAIS	0.774	0.784
≤ 31 *employees*	RAIS	0.953	0.959

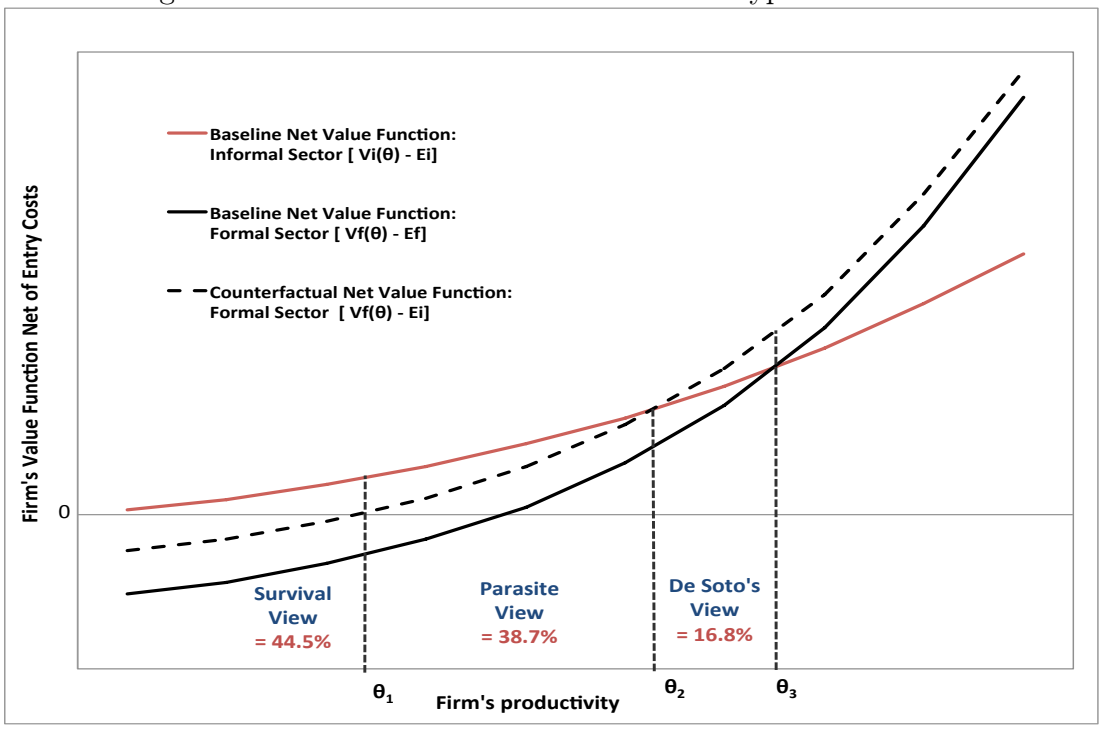
Notes: PNAD is the National Household Survey, a repeated cross-section available annually; ECINF (*Pesquisa de Economia Informal Urbana*) is a repeated cross-section of small firms (up to five employees), available only for 1997 and 2003; both data sets are collected by the Brazilian Bureau of Statistics (IBGE). RAIS (*Relacao Anual de Informacoes Sociais*) is an administrative data set collected by the Ministry of Labor, which provides an annual panel with the universe of formal firms and workers. All moments are computed using data from 2003, which is the last year ECINF is available.

6 Quantitative results

6.1 The distribution of informal firms' types in the data

In this section I use the estimation results and the taxonomy discussed in Section 4 to back out from the data the distribution of informal firms' types. For that, I use the estimated model to obtain, for each firm θ, the baseline net value function of being formal and informal: $V_f(\theta) - E_f$ and $V_i(\theta) - E_i$, respectively. I then simulate the counter-factual scenario where entry costs into the formal sector are equalized to informal sector's ($E_f = E_i$), and compute for all firms the counter-factual value of being formal once entry costs are removed, which is given by $V_f^c(\theta) - E_i$, where again the superscript c denotes the counter-factual scenario. Figure 3 revisits Figure 2 by displaying the corresponding empirical curves.

Figure 3: The distribution of informal firms types in the data



Note: The figure shows, for each productivity level, firms' value function net of entry costs in the formal and informal sectors, $V_f(\theta) - E_f$ and $V_i(\theta) - E_i$, respectively. The third curve displays the net value of being formal in a scenario where entry costs into the formal sector are equalized to informal sector's ($E_f = E_i$): $V_f^c(\theta) - E_i$, where the prime indicates the counterfactual scenario.

The relative sizes of each view are obtained by computing the mass of firms within each of the three intervals. Hence, they crucially depend on the effective productivity distribution in the informal sector, which is determined by three main elements: (i) the underlying pre-entry productivity distribution F_ν; (ii) the determinants of firms' sorting

between sectors; and (iii) the selection mechanism after entry occurs, which selects out the least productive firms.[27] As discussed in the previous section, the parameters that govern (i) and (iii) are identified by firm size distributions in both sectors and the degree of overlap between them. Element (ii) is determined by the interplay between the pre-entry productivity distribution, entry costs, and the institutional factors that determine expected profitability in both sectors, such as taxes and the costs of both margins of informality.

The data indicates that the potentially productive informal firms that are kept out of formality by high entry costs (De Soto's view) are the minority, corresponding to 16.8% of all informal firms. Parasite (type 2) firms, those that could survive as formal firms once entry costs are removed but choose to remain informal to enjoy the cost advantages of non-compliance correspond to 38.7% of all informal firms. The remaining firms, 44.5%, correspond to the survival view (type 1) firms, which are too unproductive to ever become formal and are only able to survive in the informal sector. These results thus suggest that eradicating informal firms through higher enforcement would be a more effective policy to reduce informality and improve economic performance than simply reducing entry costs. In the following section I examine the firm-level and aggregate impacts of different policies towards informality.

6.2 Firm-level and aggregate impacts of formalization policies

In this section I analyze the impacts of formalization policies at the firm level and how they aggregate up to different economy-wide effects. I consider four experiments: (i) equalizing formal and informal sectors' entry costs; (ii) a 20 p.p. cut in the payroll tax (which corresponds to eliminating social security contribution); (iii) increasing the cost of being an informal firm (the extensive margin); and (iv) increasing formal firms' costs of hiring informal workers (the intensive margin). The latter two could be achieved through greater monitoring efforts by the government, which in the model translates into lower values of the parameter b_s, $s = i, f$. In what follows I analyze the effects at the firm and aggregate levels separately.

6.2.1 The impacts on firms

To analyze the effects on firms, I follow the policy evaluation literature and use the marginal treatment effect (MTE) as impact measure [e.g. Heckman and Vytlacil (2005)]. The outcome considered here is firm's post-entry value function, $V(\theta)$, net of entry costs.

[27] The Appendix D.1 contains the derivation of post-entry productivity distributions.

For firms that remain in the same sector both in the baseline and counter-factual scenarios, the MTE is simply given by

$$\Delta(\theta; Stayers) = \log\left(V_s^c(\theta)\right) - \log\left(V_s(\theta)\right) \tag{9}$$

where again θ denotes firm's productivity, the superscript c denotes the counter-factual scenario, and $s = i, f$ indexes firm's sector (formal and informal).

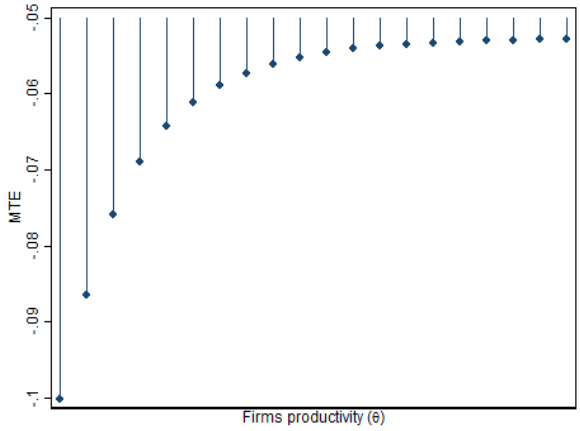
Firms that are informal in the baseline and decide to switch to the formal sector in the counter-factual scenario must pay the difference in entry costs between sectors, $\tilde{E} = E_f - E_i$. Their net, counter-factual payoff in the formal sector will be $V_f^c(\theta) - \tilde{E}$, and the MTE for switchers can thus be written as:

$$\Delta(\theta; Switchers) = \log\left(V_f^c(\theta) - \tilde{E}\right) - \log\left(V_i(\theta)\right) \tag{10}$$

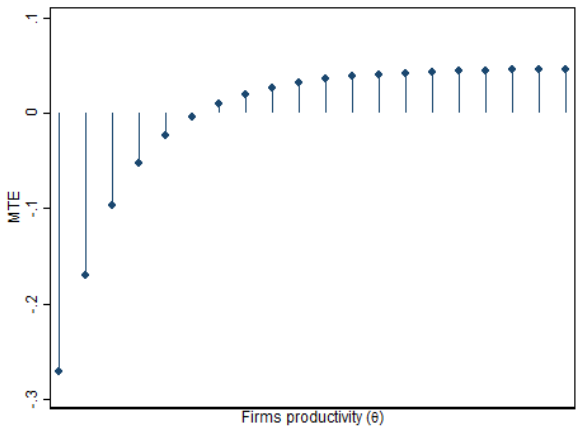
Therefore, firms can be divided into three basic groups: (i) formal stayers, which are those that are formal in the baseline and choose to remain formal in the counter-factual scenario; (ii) informal stayers, which are informal in both the baseline and counter-factual scenarios; and (iii) movers, which are the informal firms that choose to formalize after the policy is implemented. I start by analyzing the MTE-productivity profiles within each of these groups of firms and for all four policy interventions considered. For each policy and group, I compute the average MTE at the points in firms' productivity grid. Figure 4 shows the profiles for the policies that seek to reduce regulatory costs – entry costs and payroll tax – which correspond to policies (i) and (ii) above. Figure 5 shows the profiles for policies (iii) and (iv), which increase the costs of the extensive and intensive margins of informality, respectively.

Figure 4 [panel (a)] shows that all formal incumbents are hurt when formal sector's entry costs are reduced, but specially the less productive ones. This is due to the fact that reducing entry costs induces greater entry into the formal sector, which increases competition and therefore the equilibrium wage (Table 6). Because of these general equilibrium effects on wages, informal incumbents are also negatively affected by this policy [panel (c)], with a larger negative effect on average (Table 5). Turning to the payroll tax reduction, it substantially hurts low productivity formal incumbents but it has positive impacts on high productivity ones, with the MTE increasing monotonically from the lowest to the highest productivity firm [panel (b)]. This negative effect on low productivity formal firms and on all informal incumbents [panel (d)] once again comes from the general equilibrium effects on wages: as the payroll tax decreases the demand for labor increases, leading to a strong positive effect on the equilibrium wage. Due to this

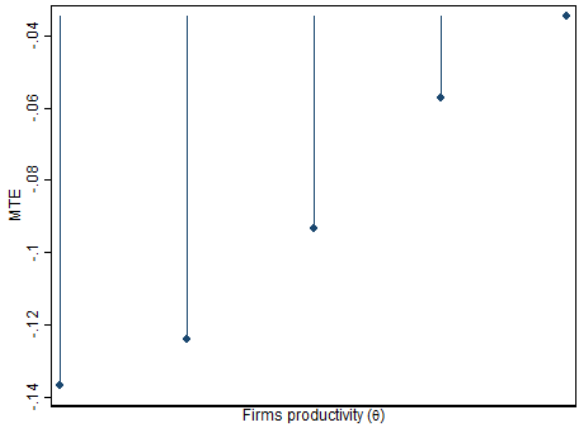
Figure 4: Profiles of Marginal Treatment Effects: Reducing regulatory costs



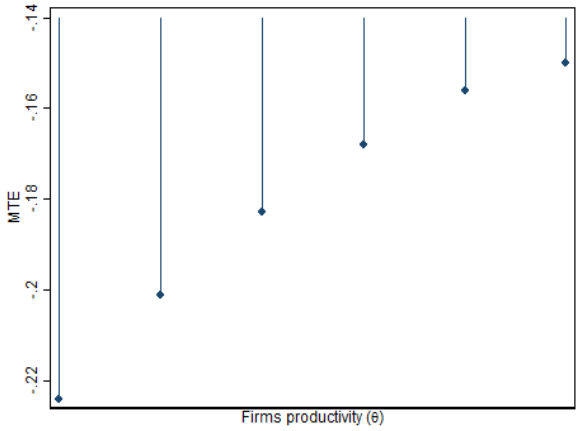
(a) Equalizing entry costs: Formal stayers



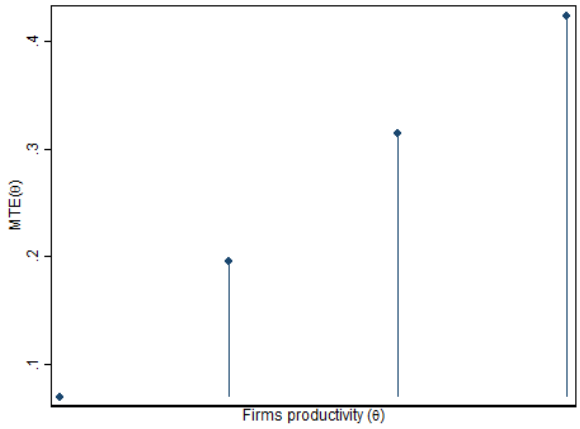
(b) Reducing Payroll Tax: Formal Stayers



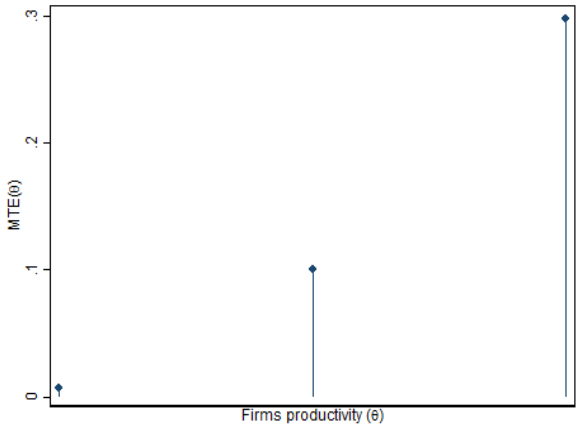
(c) Equalizing entry costs: Informal stayers



(d) Reducing Payroll Tax: Informal Stayers

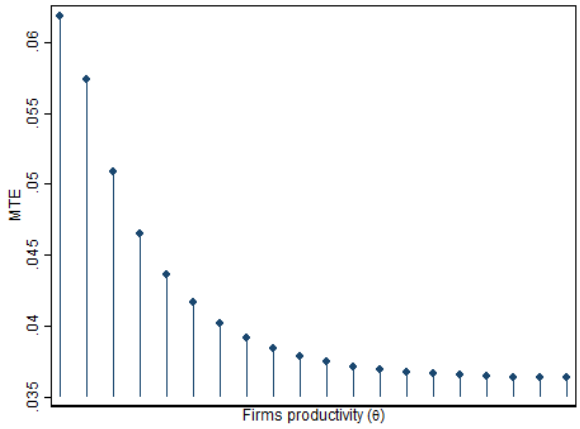


(e) Equalizing entry costs: Movers

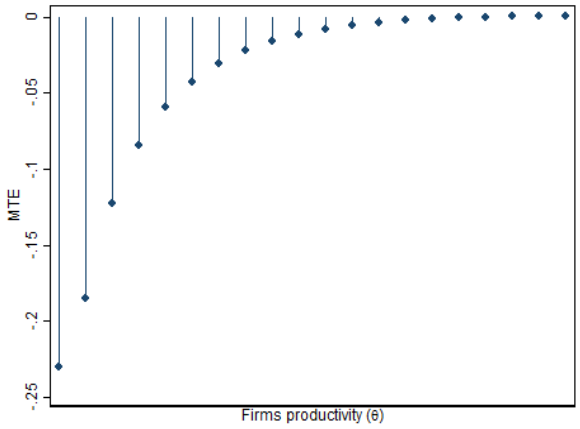


(f) Reducing Payroll Tax: Movers

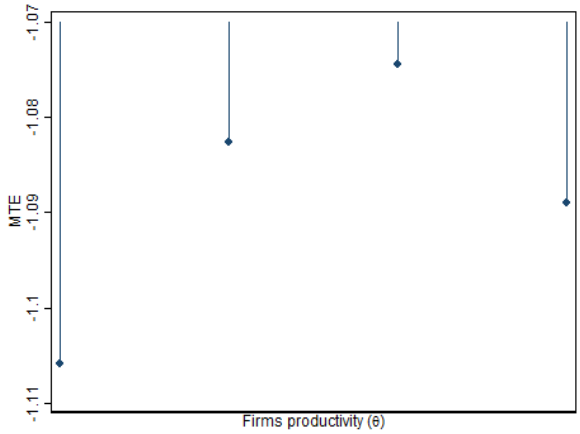
Figure 5: Profiles of Marginal Treatment Effects: Increasing the costs of informality



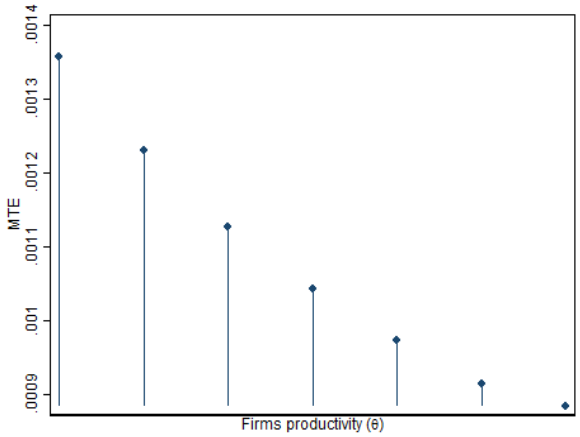
(a) Extensive Margin: Formal stayers



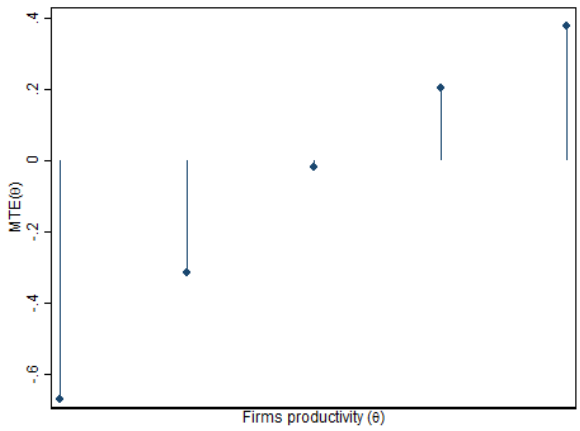
(b) Intensive Margin: Formal Stayers



(c) Extensive Margin: Informal stayers



(d) Intensive Margin: Informal Stayers



(e) Extensive Margin: Movers

wage increase, low productivity firms reduce their scale and resources are shifted away from these firms to more productive ones, which explains the increasing MTE profile observed in panel (b). Finally, informal firms that formalize due to either policy – panels (e) and (f) – greatly benefit from it, with an average increase in firms' value function of nearly 24% in the entry cost experiment and 13.2% in the payroll tax experiment (Table 5). As Table 5 shows, there is a large degree of heterogeneity in the MTEs for movers, specially in the payroll tax experiment, where the effects range from a 0.7% gain in firms' value function (the 25^{th} and 50^{th} percentiles) to a 54.9% gain in the 95^{th} percentile of the MTE distribution.

Turning to the experiments that increase the costs of informality, Figure 5 [panel (a)] shows that formal incumbents benefit from higher enforcement on the extensive margin of informality, with an average increase of 4.9% in their value functions (Table 5). Interestingly, low-productivity formal firms benefit the most, which indicates that they are the ones most directly affected by the competition from informal firms. Increasing enforcement on the extensive margin induces some informal firms to formalize and displaces a large share of informal firms. Those that survive have to greatly reduce their scale in order to stay invisible to the government, which causes them to experience a very large negative impact, with an average reduction of 66.4% in their value functions (Table 5).

Increasing the costs of the intensive margin of informality is most harmful to low productivity formal firms, as these firms hire a large fraction of their labor force without a formal contract. Thus, increasing enforcement on this margin of informality substantially increases effective labor costs for these firms. On the contrary, this policy has very limited impacts on informal firms albeit slightly positive. This result comes from general equilibrium effects, which lead to a small reduction in the equilibrium wage due to a lower demand for labor from low productivity formal firms.

6.2.2 Economy-wide effects

Reducing formal sector's entry cost leads to a substantial reduction in the share of informal firms, of nearly 22 p.p. (Table 6). The effect on the share of informal workers is however nearly null, which highlights the importance of accounting for the intensive margin of informality: the share of formal firms grows due to the formalization of low-productivity firms, which hire a large share of their labor force without a formal contract, and therefore the net effect on *labor informality* is near zero.

The opposite is true when the payroll tax is reduced: informal employment is substantially reduced but the share of informal firms does not fall as much. This is observed because the labor tax directly affects formal firms' decision to hire informal or formal la-

Table 5: The Distribution of Marginal Treatment Effects

	All Firms	Formal Stayers	Informal Stayers	Movers
		Reducing Entry Costs		
Mean	-0.088	-0.071	-0.105	0.215
Pctile 25	-0.054	-0.076	-0.045	0.120
Pctile 50	-0.043	-0.064	-0.043	0.192
Pctile 75	-0.038	-0.058	-0.038	0.324
Pctile 95	-0.033	-0.054	-0.035	0.448
		Reducing Payroll Tax		
Mean	-0.173	-0.045	-0.197	0.124
Pctile 25	-0.212	-0.095	-0.212	0.007
Pctile 50	-0.191	-0.023	-0.200	0.007
Pctile 75	-0.167	0.015	-0.174	0.246
Pctile 95	0.007	0.038	-0.155	0.438
		Higher Enforcement - Extensive Mg.		
Mean	-0.793	0.048	-1.091	-0.165
Pctile 25	-1.092	0.040	-1.105	-0.510
Pctile 50	-1.074	0.044	-1.092	-0.143
Pctile 75	-0.306	0.051	-1.074	0.112
Pctile 95	0.068	0.076	-1.066	0.399
		Higher Enforcement - Intensive Mg.		
Mean	-0.009	-0.072	0.001	–
Pctile 25	0.001	-0.100	0.001	–
Pctile 50	0.001	-0.059	0.001	–
Pctile 75	0.001	-0.026	0.001	–
Pctile 95	0.001	-0.006	0.001	–

Note: The MTEs are computed used the expressions (9) and (10), defined in the text.

bor; however, firms' formalization is also heavily influenced by formal sector's entry cost, which remains unaltered. Increasing enforcement on the intensive margin is the least effective policy to reduce informality, as it only generates a small reduction in the share of informal workers and leads to greater informality among firms. The latter effect is observed because the effective cost of being formal increases for less productive firms, as it is now harder for formal firms to hire informal workers, which increases their incentives to become informal.

These subtler policy impacts can only be unveiled if one explicitly considers the intensive margin. The existing literature has focused on the extensive margin alone, and therefore reducing firm informality necessarily leads to lower labor informality. As the above results show, this is no longer the case if one accounts for the intensive margin, and

firm and labor informality can actually move in opposite directions as firms optimally respond to different policies towards informality.

Table 6: Main aggregate outcomes

	Baseline	Entry Costs	Payroll Tax	Extensive Mg.	Intensive Mg.
Informal workers (share)	0.352	0.351	0.228	0.138	0.301
Informal firms (share)	0.695	0.477	0.613	0.039	0.728
Informal GDP (share)	0.260	0.198	0.191	0.005	0.271
GDP	1.000	1.035	0.991	0.994	0.984
TFP	1.000	0.951	1.076	1.133	1.005
Olley & Pakes	0.559	0.657	0.597	0.684	0.536
Wages	1.000	1.029	1.141	0.981	0.999
Taxes	1.000	1.094	0.881	1.216	0.986
Mass of active firms	1.000	1.210	0.851	0.651	0.975
Welfare	1.000	1.199	0.973	1.012	0.987

Notes: The variation in average log-TFP is measured as $\exp\{\log(TFP)_c - \log(TFP)_b\}$, where $\log(TFP)_c$ and $\log(TFP)_b$ denote the log-TFP in the counterfactual and baseline scenarios, respectively. The Olley & Pakes indicator corresponds to a slightly modified version of the indicator proposed by Olley and Pakes (1996). It is obtained through a simple decomposition of the weighted average firm-level TFP: $\hat{\Theta} = s_I\left[\bar{\theta}_I + \sum_{j\in I}(\theta_j - \bar{\theta}_I)(s_{j,I} - \bar{s}_I)\right] + (1-s_I)\left[\bar{\theta}_F + \sum_{j\in F}(\theta_j - \bar{\theta}_F)(s_{j,F} - \bar{s}_F)\right]$, where $\bar{\theta}_k$ is the unweighted average of firm-level productivity in sector $k = I, F$, and $\bar{s}_k$ is the average share. The indicator within each sector is the covariance term between brackets, while the economy wide indicator corresponds to the weighted average of both sector specific indicators.

Reducing entry costs also eliminates dead weight losses from wasteful barriers to entry, which substantially increases the mass of active firms in the economy (21% larger relatively to the baseline). As a result, competition, production in the formal sector, GDP and wages increase. In addition to greater firm entry, a second channel that causes GDP to increase is the fact that newly formalized firms are no longer size constrained as they would be in the informal sector. Since there are no incentives to remain inefficiently small to avoid being detected by the government, these firms' production increases in the formal sector. However, the intervention has a negative effect on aggregate TFP because more low-productivity firms enter the formal sector and are now more likely to survive (as formal sector's exit rate is lower), which has a negative composition effect on TFP.

Increasing enforcement on the extensive margin is highly effective in reducing informality measured as the share of firms, workers or GDP. This generates a large positive effect on aggregate TFP (a 13.3% increase) due to composition effects, as it eliminates a large fraction of low-productivity firms, which results in a mass of active firms 35% lower

than in the baseline. The GDP remains roughly unchanged, which is a consequence of these two opposing effects: higher TFP going in the direction of increasing production and lower mass of firms in the opposite direction.

The welfare analysis shows that reducing entry costs leads to a substantial welfare gain of nearly 20%. This result is a consequence of the substantial increase in the mass of active firms, higher GDP, wages and tax revenues; additionally, this policy mechanically increases aggregate net profits (which enters directly the welfare measure), as formal sector's entry costs are substantially reduced. Higher enforcement on the extensive margin also has a positive but smaller effect on welfare (a 1.2% increase), which is a result of two counteracting forces. On the one hand, it leads to a substantial increase in tax revenues (21.6%), which is rebated directly to the households.[28] On the other hand, this policy eradicates almost all informal firms, substantially reducing the mass of active firms, and reduces the equilibrium wage. It is worth highlighting that this should be seen as an upper bound for the impacts of greater enforcement on welfare, as the experiment makes two strong assumptions: (i) all tax revenues are directly rebated to households, with no resources lost; and (ii) there is no cost of implementing greater enforcement. The latter is likely to be substantial, since monitoring a large number of small firms is likely to be costly.

Finally, even though the different interventions always manage to reduce at least one measure of informality, they do not always lead to welfare improvements. This is the case of higher enforcement on the intensive margin and lower payroll tax. In the first case, the policy reduces the share of informal workers but negatively affects small formal firms and actually leads to an increase in the share of informal firms. These effects, combined with a slight decline in tax revenues, cause welfare to decrease. As for the lower payroll tax intervention, its general equilibrium effects (i.e. wage increases) undo some its firm-level benefits. More importantly, there is a substantial reduction in tax revenues, which has direct negative impacts on welfare.

7 Final remarks

This paper investigates the role of informal firms in economic development, how they respond to different formalization policies and their effects on overall economic performance. I develop a framework that distinguishes between two margins of informality: (i) when firms do not register and pay entry fees (extensive margin); and (ii) when firms pay workers "off the books" (intensive margin). The latter is a central innovation, as it

[28] This positive effect can be interpreted as a stylized version of the mechanisms highlighted by the literature on fiscal capacity [e.g. Besley and Persson (2013)].

is empirically important and allows to unveil new and non-obvious firm-level responses to policy changes regarding informality decisions. Accounting for the intensive margin also has direct implications to our understanding of informality, as it breaks the direct association between worker and firm informality. In particular, formal and informal are no longer disjoint states for firms, as formal firms may hire part or all of their labor force informally.

The framework developed here integrates the leading views of informality in an unified setting, and provides a natural taxonomy of informal firms based on these views. I take the model to data on formal and informal firms in Brazil to back out the empirical relevance of these competing views. The results show that firms that are potentially productive and which formalize and succeed when formal sector's entry costs are removed constitute a small fraction of all informal firms (16.8%). The view that argues that informal firms choose informality to exploit the cost advantages of non-compliance even though they are productive enough to survive in the formal sector corresponds to a large fraction of all informal firms, 38.7%. The remaining firms correspond to those too unproductive to ever become formal.

Counterfactual analysis of policy effects shows that no single policy generates positive effects for all firms and that there is substantial heterogeneity in policy effects between groups (i.e. informal to formal switchers, formal stayers and informal stayers) and within groups. At the aggregate level, I find that increasing enforcement is highly effective in reducing informality but it does not increase GDP and barely increases welfare (an upper bound effect of 1.2%). Reducing formal sector's entry costs is not as effective in reducing informality but generates substantial welfare gains and leads to greater GDP and wages. Overall, the results show that informality reductions can be but are not necessarily associated to higher GDP, TFP or welfare.

References

Abbring, J. H. (2010). Identification of dynamic discrete choice models. *Annual Review of Economics 2*(1), 367–394.

Ackerberg, D., C. L. Benkard, S. Berry, and A. Pakes (2007). Econometric tools for analyzing market outcomes. *Handbook of econometrics 6*, 4171–4276.

Adda, J. and R. Cooper (2003). *Dynamic Economics: Quantitative methods and applications*. The MIT Press.

Aguirregabiria, V. and P. Mira (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics 156*(1), 38–67.

Almeida, R. and P. Carneiro (2009). Enforcement of labor regulation and firm size. *Journal of Comparative Economics 37*(1), 28 – 46.

Almeida, R. and P. Carneiro (2012). Enforcement of labor regulation and informality. *American Economic Journal: Applied Economics 4*(3), 64–89.

Amaral, P. S. and E. Quintin (2006). A competitive model of the informal sector. *Journal of Monetary Economics 53*(7), 1541–1553.

Arias, J., O. Azuara, P. Bernal, J. Heckman, and C. Villarreal (2010). Policies to promote growth and economic efficiency in mexico. NBER Working Paper 16554.

Besley, T. and T. Persson (2013). Taxation and development. Forthcoming.

Bosch, M., E. Goñi-Pacchioni, and W. Maloney (2012). Trade liberalization, labor reforms and formal–informal employment dynamics. *Labour Economics 19*(5), 653–667.

Bruhn, M. (2011). License to sell: The effect of business registration reform on entrepreneurial activity in mexico. *Review of Economics and Statistics 93*(1), 382–386.

Charlot, O., F. Malherbet, and C. Terra (2011). Product market regulation, firm size, unemployment and informality in developing economies. IZA Discussion Papers No.5519.

Cosar, A., N. Guner, and J. Tybout (2014). Firm dynamics, job turnover, and wage distributions in an open economy. Mimeo.

de Andrade, G., M. Bruhn, and D. McKenzie (2013). A helping hand or the long arm of the law ? experimental evidence on what governments can do to formalize firms. Policy Research Working Paper, WPS6435.

De Mel, S., D. McKenzie, and C. Woodruff (2013). The demand for, and consequences of, formalization among informal firms in sri lanka. *American Economic Journal: Applied Economics 5*(2), 122–150.

de Paula, A. and J. A. Scheinkman (2010). Value-added taxes, chain effects, and informality. *American Economic Journal: Macroeconomics 2*(4), 195–221.

de Paula, A. and J. A. Scheinkman (2011). The informal sector: An equilibrium model and some empirical evidence. *Review of Income and Wealth 57*, S8–S26.

De Soto, H. (1989). *The Other Path*. Harper e Row, New York.

D'Erasmo, P. and H. Boedo (2012). Financial structure, informality and development. *Journal of Monetary Economics 59*(3), pp. 286–302.

Dix-Carneiro, R. (2014). Trade liberalization and labor market dynamics. *Econometrica 82*(3), 825–885.

Dix-Carneiro, R. and B. Kovak (2014). Trade reform and regional dynamics: Evidence from 25 years of brazilian matched employer-employee data. Mimeo.

Fajnzylber, P., W. F. Maloney, and G. V. Montes-Rojas (2011). Does formality improve micro-firm performance? evidence from the brazilian simples program. *Journal of Development Economics 94*(2), 262 – 276.

Farrell, D. (2004). The hidden dangers of the informal economy. *McKinsey Quarterly*, 26–37.

Fortin, B., N. Marceau, and L. Savard (1997). Taxation, wage controls and the informal sector. *Journal of Public Economics 66*(2), 293–312.

Galiani, S. and F. Weinschelbaum (2012). Modeling informality formally: households and firms. *Economic Inquiry 50*(3), 821–838.

Goldberg, P. K. and N. Pavcnik (2003). The response of the informal sector to trade liberalization. *Journal of Development Economics 72*(2), 463–496.

Goldberg, P. K. and N. Pavcnik (2007). Distributional effects of globalization in developing countries. *Journal of economic literature 45*(1), 39–82.

Gourieroux, C. and A. Monfort (1996). *Simulation-Based Econometric Methods*. Oxford University Press.

Gourinchas, P.-O. and J. Parker (2002). Consumption over the life cycle. *Econometrica 70*(1), 47–89.

Heckman, J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrics 73*(3), 669–738.

Heckman, J. J. (2001). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy 109*(4), pp. 673–748.

Heckman, J. J. and S. Navarro (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics 136*(2), 341–396.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics 124*(4), 1403 – 1448.

Hsieh, C.-T. and B. A. Olken (2014). The missing" missing middle". NBER Working Paper No. 19966.

Ichimura, H. and C. Taber (2002). Semiparametric reduced-form estimation of tuition subsidies. *The American Economic Review 92*(2), pp. 286–292.

Kaplan, D. S., E. Piedra, and E. Seira (2011). Entry regulation and business start-ups: Evidence from mexico. *Journal of Public Economics 95*(11), 1501–1515.

Keane, M. P., P. E. Todd, and K. Wolpin (2011). The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications. Volume 4A of *Handbook of Labor Economics*, Chapter 04, pp. 331–461. Elsevier.

La Porta, R. and A. Shleifer (2008). The unofficial economy and economic development. *Brooking Papers on Economic Activity 105*(3), 473–522.

La Porta, R. and A. Shleifer (2014). Informality and development. Forthcoming.

Leal Ordonez, J. C. (2014). Tax collection, the informal sector, and productivity. *Review of Economic Dynamics 17*(2), 262–286.

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

Luttmer, E. (2007). Selection, growth and the size distribution of firms. *The Quarterly Journal of Economics 122*(3), 1103–1144.

Matzkin, R. L. (1992). Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica 60*(2), pp. 239–270.

Matzkin, R. L. (2003, 09). Nonparametric estimation of nonadditive random functions. *Econometrica 71*(5), 1339–1375.

Meghir, C., R. Narita, and J. Robin (2014). Wages and informality in developing countries. Mimeo.

Menezes-Filho, N. A. and M.-A. Muendler (2011). Labor reallocation in response to trade reform. NBER Working Paper No. 17372.

Monteiro, J. and J. J. Assunção (2012). Coming out of the shadows? estimating the impact of bureaucracy simplification and tax cut on formality in brazilian microenterprises. *Journal of Development Economics 99*(1), 105–115.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Perry, G., W. Maloney, O. Arias, P. Fajnzylber, A. Mason, and J. Saavedra-Chanduvi (2007). *Informality: Exit or exclusion*. The World Bank. Washington, D.C.

Prado, M. (2011). Government policy in the formal and informal sectors. *European Economic Review 55*, 1120–1136.

Rauch, J. E. (1991). Modeling the informal sector formally. *Journal of Development Economics 35*(1), 33–47.

Rocha, R., L. Rachter, and G. Ulyssea (2014). Do entry regulation and taxes hinder firm creation and formalization? evidence from brazil. Mimeo.

Rust, J. (1994). Structural estimation of markov decision processes. Volume 4 of *Handbook of Econometrics*, pp. 3081 – 3143. Elsevier.

Schneider, F. (2005). Shadow economies around the world: what do we really know? *European Journal of Political Economy 21*(3), 598 – 642.

Taber, C. R. (2000). Semiparametric identification and heterogeneity in discrete choice dynamic programming models. *Journal of econometrics 96*(2), 201–229.

Tauchen, G. (1986). Finite state markov-chain approximation to univariate and vector autoregressions. *Economic Letters 20*, 177–81.

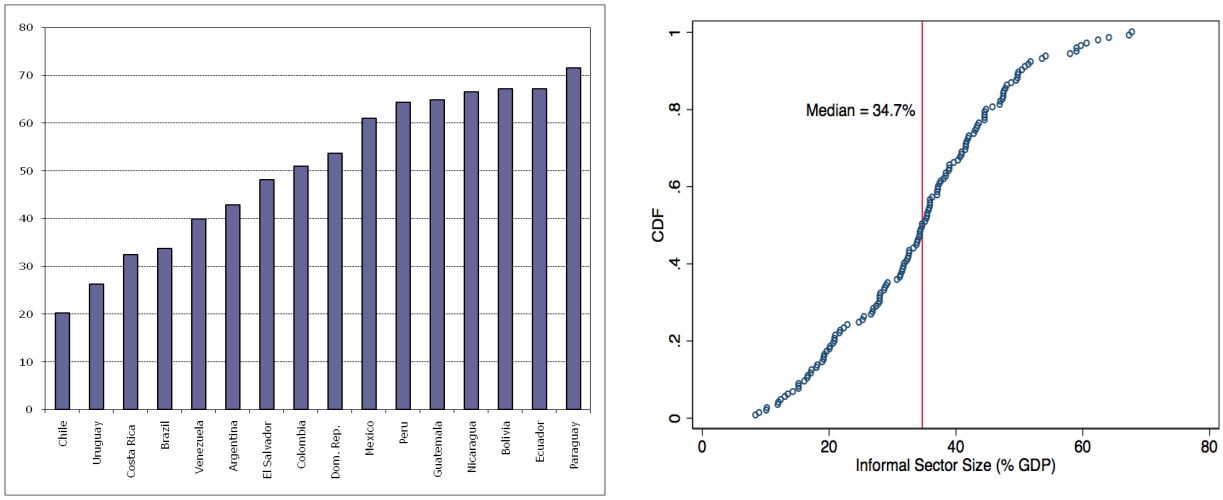
Ulyssea, G. (2010, January). Regulation of entry, labor market institutions and the informal sector. *Journal of Development Economics 91*, 87–99.

Ulyssea, G. (2014, February). Firms dynamics in developing countries. Unpublished manuscript.

A Cross country evidence

Figure A.1's left panel displays informal sector's size in Latin American countries, which is measured as the share of employees not covered by social security. [29] The right panel shows the c.d.f. plot of informal sector's size for 116 countries that have a GDP per capita that is less or equal to half of USA's. The size measure used in this graph is informal sector's share of GDP, which comes from La Porta and Shleifer (2008)'s data set that uses Schneider (2005)'s methodology.

Figure A.1: Informal sector's size



(a) Labor informality (Latin America)

(b) CDF of size measure (developing countries)

B Data appendix

As described in Section 2, the two main data sets used in this paper are the ECINF survey (*Pesquisa de Economia Informal Urbana*) and the *Relacao Anual de Informacoes Sociais* (RAIS), an administrative data set from the Brazilian Ministry of Labor. In both data sets, I only keep firms that operate in manufacturing, services or commerce, thus excluding public sector and agriculture. In RAIS, I exclude firms that declare a wage bill equal to zero. Since the RAIS data set contains the universe of formal firms, I use a 25% random sample from the original data set to decrease the computational burden.

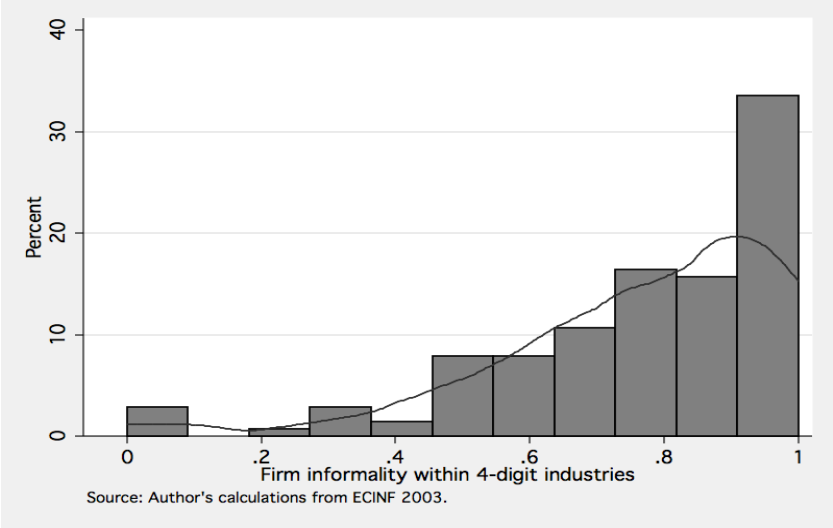
As for the ECINF, some additional filters were applied. Many of the observations regard self employed individuals, street vendors and other activities that do not correspond to the standard definition of a firm. In order to obtain the most comparable unit of analysis with the formal firms covered by the RAIS data set, I dropped the entrepreneurs who declared to have another job, and who do not have a specific physical location outside

[29] The data come from the Socio-Economic Data Base for Latin America and the Caribbean (SEDLAC), a joint initiative by the World Bank and Universidad Nacional de La Plata (available at *http://sedlac.econo.unlp.edu.ar/esp/*).

their household where their activity takes place. To avoid outliers, I trimmed the first and 99th percentiles of log-revenues distribution.

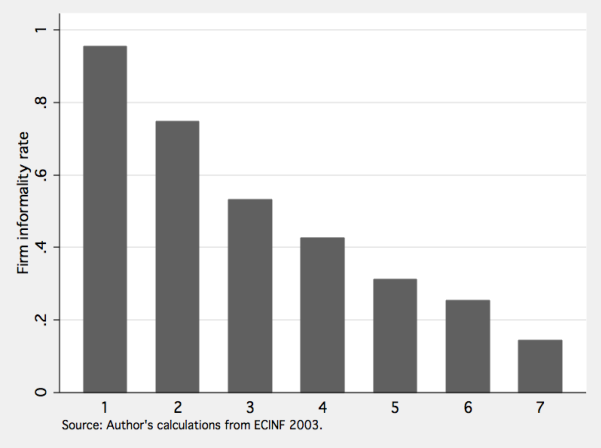
C Additional Stylized facts

Figure B.1: Share of informal firms at the 4-digit industry level: Histogram

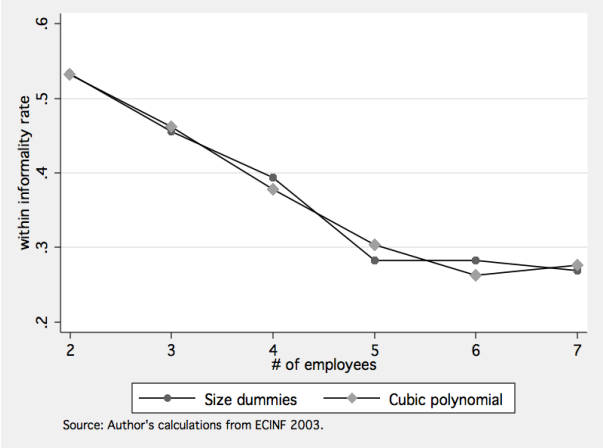


Note: The variable used is the share of informal firms measured at the 5-digit industry level. The figure shows the histogram of this industry-specific measure of firm informality.

Figure B.2: Informality margins and firms' size



(a) Extensive margin



(b) Intensive margin

Note: The panel on the left shows the share of informal firms among firms with size $n = 1, \ldots, 7$ (where size is measured as number of employees). The panel on the right shows the average share of informal workers within formal firms, among firms with size $n = 2, \ldots, 7$.

D Model Appendix

D.1 Productivity distributions in both sectors

The post-entry, unconditional productivity distribution in the informal and formal sectors, respectively, is given by the following expressions:

$$\begin{aligned} f_{\theta_i}(x) &= \frac{1}{G(\overline{\nu}_f) - G(\overline{\nu}_i)} \int_{\overline{\nu}_i}^{\overline{\nu}_f} f(x|\nu)\, dG(\nu) \\ f_{\theta_f}(x) &= \frac{1}{1 - G(\overline{\nu}_f)} \int_{\overline{\nu}_f}^{\infty} f(x|\nu)\, dG(\nu) \end{aligned} \tag{11}$$

where f_{θ_s} is absolutely continuous and $F_{\theta_s}(\cdot)$ denotes the corresponding c.d.f..

As mentioned above, firms can be surprised with a bad productivity draw. Those with a $\theta < \overline{\theta}_s$, where $\overline{\theta}_s$ is such that $\pi_s\left(\overline{\theta}_s, w\right) = 0$, will not produce and will leave immediately. Hence, the effective productivity distribution among successful entrants is given by the following expressions:

$$\tilde{f}_{\theta_s}(x) = \begin{cases} \frac{f_{\theta_s}(x)}{1 - F_{\theta_s}\left(\overline{\theta}_s\right)} & \text{if} \quad x \geq \overline{\theta}_s \\ 0 & \text{if} \quad \theta < \overline{\theta}_s \end{cases} \tag{12}$$

where $s = i, f$.

D.2 Uniqueness of Equilibrium

This section contains a simple argument to prove the uniqueness of equilibrium. The key equilibrium conditions are given by the zero profit conditions, the free entry conditions and the market clearing condition, respectively:

$$\pi_s\left(\bar{\theta}_s, w\right) \equiv \Pi_s\left(\bar{\theta}_s, w\right) - \bar{c} = 0 \tag{13}$$

$$V_i^e(\overline{\nu}_i, w) = c_i \tag{14}$$

$$V_f^e(\overline{\nu}_f, w) = V_i^e(\overline{\nu}_f, w) - (c_i - c_f) \tag{15}$$

$$L_i + L_f = \bar{L} \tag{16}$$

where $s = i, f$ and the free entry conditions assume that entry is positive in both sectors.

Fix a given wage. Given the assumptions made for the cost and production functions, the functions $\pi_s(\theta, w)$ are strictly increasing in θ and decreasing in w. Moreover, as $\bar{c} > 0$, there is a $\theta > 0$ such $\pi_s(\theta, w) < 0$. Thus, there is a unique $\bar{\theta}_s$ such that 13 holds. The simple form of the value functions, $V_s(\theta, w) = \max\left\{0, \frac{\pi_s(\theta, w)}{\kappa_s}\right\}$, implies that they are also continuous and strictly increasing in θ and decreasing in w. Combining this last fact with the assumptions made about $F(\theta|\nu)$, it follows that that there is an unique $\bar{\nu}_s$, $s = i, f$, such that free entry conditions hold, and that $\bar{\nu}_f > \bar{\nu}_i$. The latter follows from the assumption that $c_f > c_i$. The unique entry thresholds pin down the mass of

entrants in both sectors: $M_i = [G(\overline{\nu}_f) - G(\overline{\nu}_i)]\, M$ and $M_f = [1 - G(\overline{\nu}_f)]\, M$. Given the mass of entrants in each sector, and the unique thresholds $\bar{\theta}_s$, the flow conditions in both sectors [given by (6)] pin down the mass of firms in each sector, μ_s. The last condition to close the equilibrium determination is the market clearing condition for the labor market, which determines the equilibrium wage. Of course, if $L^d \equiv L_i + L_f > \bar{L}$ there is excess demand and the wage will increase up until the point where $L^d = \bar{L}$ (the symmetric argument is true for excess supply). Because of the properties of the profit functions, the individual labor demand functions $\ell^*(\theta, w)$ are also continuous, single valued, and strictly increasing in θ and decreasing in w. Thus, there is an unique wage such that $L^d = \bar{L}$.

□

E Estimation Appendix

E.1 Standard Errors

The second-stage, SMM estimator is given by

$$\hat{\varphi} \;=\; \arg\min_{\varphi} Q(\varphi) = \left\{ g(\varphi)' \hat{\mathbf{W}} g(\varphi) \right\}$$

where $g(\varphi) = \hat{m} - m^s(\varphi)$ and I omit the conditioning arguments for notational convenience.

The following assumptions are made for asymptotic normality to hold: (i) φ_0 and $\hat{\varphi}$ are interior to the parameter space; (ii) the simulator used to generate the simulated data is continuously differentiable w.r.t. φ in a neighborhood $\mathcal{B}$ of φ_0; and (iii) $\mathbf{G}_0 \equiv E[\nabla_\varphi g(\varphi_0)]$ exists, is finite and $\mathbf{G}_0'\mathbf{W}\mathbf{G}_0$ is nonsingular.

The first order condition of the SMM estimator is given by $\nabla_\varphi g_s(\hat{\varphi})' \hat{\mathbf{W}} g_s(\hat{\varphi}) = 0$, where $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$, and $\mathbf{W}$ is positive semi-definite. Using the mean value theorem to expand $g_s(\hat{\varphi})$ around φ_0 and combining with the FOC gives

$$\sqrt{N}(\hat{\varphi} - \varphi_0) = -\left[\nabla_\varphi g_s(\hat{\varphi})' \hat{\mathbf{W}} \nabla_\varphi g_s(\overline{\varphi}) \right]^{-1} \nabla_\varphi g_s(\hat{\varphi})' \hat{\mathbf{W}} \sqrt{N} g_s(\varphi_0)$$

Given that the simulator is unbiased and by the CLT, $\sqrt{N} g_s(\varphi_0)$ converges to a Normal distribution with zero mean and the following asymptotic variance

$$\Sigma_s = \left(1 + \frac{1}{S}\right) \Sigma$$

where $\Sigma = E\left[g(\varphi_0)\, g(\varphi_0)' \right]$ is the GMM asymptotic variance-covariance matrix.

Finally, from the WLLN both $N^{-1}\nabla_\varphi g_s(\hat{\varphi})$ and $N^{-1}\nabla_\varphi g_s(\overline{\varphi})$ converge in probability to G_0 and by the Slutzky theorem one gets

$$\sqrt{N}(\hat{\varphi} - \varphi_0) \xrightarrow{d} N\left(0, (\mathbf{G}_0'\mathbf{W}\mathbf{G}_0)^{-1} \mathbf{G}_0'\mathbf{W}\Sigma_s\mathbf{W}\mathbf{G}_0 (\mathbf{G}_0'\mathbf{W}\mathbf{G}_0)^{-1}\right)$$

Analogous to the GMM estimator, the optimal weighting matrix is given by $W^* = \Sigma_s^{-1}$, which therefore reduces the asymptotic variance-covariance matrix to

$$V_s(W^*) = \left(\mathbf{G}_0'\mathbf{\Sigma_s^{-1}}\mathbf{G}_0\right)^{-1}$$

The actual variance-covariance is computed using the empirical counterpart of Σ (estimated from real data), and the computational equivalent of G_0, which can be obtained using standard numerical differentiation methods.

E.2 Simulation Algorithm

For the simulations, I consider a mass of $M = 300,000$ potential entrants. For each potential entrant, I draw a pre-entry productivity parameter (ν) and a post entry productivity shock (ϵ). I use 77 equally spaced grid points for the productivity space. The maximum value in the productivity grid implies a firm's size of more than 18,000 employees and is not binding. Since each potential entrant has an individual pre-entry productivity parameter, it is necessary to compute a vector of transition probabilities for each point in the grid in order to compute the expected post-entry values in each sector for each potential entrant. For that, I use the method proposed by Tauchen (1986).

The stochastic components of the model are drawn only once in the beginning of the procedure and are kept fixed during the algorithm's execution. The estimation procedure is done conditional on the observed wage, which is the mean real wage for prime age males in the period 1997–2003 (pooling formal and informal employees together). I take the mean using the six years prior the baseline year to approximate the steady state wage. The steps of the estimation algorithm are the following:

1. Compute the observed wage.
2. Draw 300,000 observations of the random variables $X_1 \sim U(0,1)$ and $X_2 \sim N(0,1)$.
3. Compute the moments from the data, $\hat{m}$.
4. Initiate the optimization algorithm:
   (i) Guess φ.
   (ii) Obtain the post-entry productivity as $\log(\theta) = \log(\nu) + \epsilon$, where $\nu = G_\nu^{-1}(X_1)$ and $\log(\varepsilon) \equiv \epsilon = \sigma X_2$.
   (iii) Use the model to generate a simulated data set and compute $m^s(\varphi;\psi)$.
   (iv) Compute the loss function $Q(\varphi;\psi)$, as defined in (8).
   (v) Check if the objective function is minimized (according some tolerance level). If not, return to step (i) and guess a new φ.

For Online Publication

Discussion of alternative methods to identify the model non parametrically

The present framework is a very simplified version of a discrete choice dynamic programming (DCDP) model.[30] Recent developments in the literature [e.g. Heckman and Navarro (2007)] have shown that semiparametric identification is possible for some classes of dynamic discrete choice models [see Abbring (2010)]. However, in what follows I argue that non-parametric estimation of the different objects in the model is either not feasible given the goals of this paper and the data available, or the assumptions needed are not attainable.

Starting by the unconditional productivity distribution, $F(\theta)$, there are a number of approaches to estimate it non-parametrically, as long as one is willing to assume a functional form for $q(\cdot)$ (Cobb-Douglas, say). One of the most well-known approaches is the one proposed by Olley and Pakes (1996), which requires access to a panel of firms that contains information on inputs, investment and revenues or, ideally, physical production [see Ackerberg et al. (2007) for a review of more recent methods]. The data requirements are however high, and are not met in the present context.

If knowledge of the profit function alone was sufficient, one could use Matzkin (2003) to identify it nonparametrically, as profit functions are homogeneous of degree one and thus satisfy Matzkin's conditions.[31] However, the profit function is a reduced-form object and not a primitive of the model as defined by the elements in ω. In the present application, it is necessary to identify the cost and revenue functions separately in order to perform the counterfactual analysis. The cost function could be nonparametrically estimated from data on variable costs (when available) and inputs, but this would only give the relationship between inputs and costs for a given structure. Once there are changes in the intensity of government inspections say, the structure that generated the estimated cost function would no longer be valid.

Finally, Heckman and Navarro (2007) extend Taber (2000) analysis to a general finite horizon model with a rich dynamics for the unobserved shocks and are able to semiparametrically identify their full structural model (including the cost and earnings functions). Their framework however, does not apply to the family of models considered here, as they rely on additive separability between observable and unobservable state variables, and on the independence between both.[32] Additionally, their identification proof for the full structural model relies on "identification at infinity" type of arguments, which require strong support assumptions that are most likely not satisfied in the present application.

[30] The DCDP literature is quite extensive. The interested reader can refer to the well-known review of Rust (1994). More recently, Aguirregabiria and Mira (2010) and Keane et al. (2011) provide comprehensive and updated reviews of estimation methods and applications. Abbring (2010) presents a recent discussion on identification of different DCDP models.

[31] Note, however, that even in this case one would have to assume that θ is independent of wages, which is only true when firms are truly price takers.

[32] They apply results from Matzkin (1992) on nonparametric identification of static binary choice models.