American Economic Review 2018, 108(8): 2015–2047
https://doi.org/10.1257/aer.20141745

Firms, Informality, and Development: Theory and Evidence from Brazil†

By Gabriel Ulyssea*

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, 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 with higher output, TFP, or welfare. (*JEL* D22, E26, H26, J46, O14, O17)

The informal sector is a prominent feature of most developing economies,[1] which is likely to have deep economic implications. First, high informality levels imply widespread tax evasion, 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. Third, it allows less productive (informal) firms to compete with more productive (formal) firms, leading to misallocation of resources and potentially large total factor productivity (TFP) losses (e.g., Hsieh and Klenow 2009). In contrast, informality can be beneficial to growth as it provides de facto flexibility for firms that would be otherwise constrained by burdensome regulations.[2] Therefore, understanding how the informal sector affects the economy

* Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford, OX1 3UQ (email: gabriel.ulyssea@economics.ox.ac.uk). This paper was accepted to the *AER* under the guidance of Hilary Hoynes, Coeditor. I am indebted to James Heckman, Steven Durlauf, and Chang-Tai Hsieh for their guidance and constant encouragement. I would like to thank four anonymous referees for extremely helpful comments and suggestions. I am also thankful to 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, and participants at various conferences and seminars for helpful comments and discussions. Special thanks to Claudio Ferraz, Rafael Dix-Carneiro, Stephane Wolton, and Dimitri Szerman for their detailed and extremely helpful comments and feedback on previous versions of the manuscript. Of course, all errors are mine. Financial support from CAPES, IPEA, and the University of Chicago is gratefully acknowledged. The author declares that he has no relevant or material financial interests that relate to the research described in this paper.

† Go to https://doi.org/10.1257/aer.20141745 to visit the article page for additional materials and author disclosure statement.

[1] In Brazil, nearly two-thirds of businesses, 40 percent of GDP, and 35 percent of employees are informal. Similarly, the informal sector accounts for 50 percent of workers and 41.9 percent of GDP in Colombia, and 60 percent of workers and 31.9 percent of GDP in Mexico (online Appendix A contains evidence of informal sector's size for 116 countries).

[2] Meghir, Narita, and Robin (2015) develop an equilibrium wage-posting model that incorporates some of these trade-offs. Their analysis focuses on the low-skill labor market, and their results indicate that the negative effects dominate: reducing informality through greater enforcement increases welfare in the economy.

and evaluating the firm-level and aggregate impacts of policies toward informality are central questions in economic development.

This paper sheds light on these issues by developing a new framework that distinguishes two margins of informality: (i) whether firms register and pay entry fees to achieve a formal status, the *extensive 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 informality 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 (e.g., Rauch 1991; Fortin, Marceau, and Savard 1997; Amaral and Quintin 2006; de Paula and Scheinkman 2010, 2011; Bosch and Esteban-Pretel 2012). The intensive margin breaks this direct association between firm and worker informality,[3] and allows to uncover new and subtler firm-level responses to policy changes. I show that these responses translate into non-obvious and quantitatively important effects on TFP, total output, and overall informality. Empirically, I present evidence that the intensive margin accounts for a large share of total informal employment.[4]

The model is similar in spirit to Melitz (2003) but with some important innovations in the entry and production structures, as well as the addition of the two margins of informality and worker heterogeneity. Firms are ex ante heterogeneous (before entry occurs) and must decide whether to enter the formal or informal sectors. Sector membership is defined by the extensive margin, and the (in)formal sector is formed by (un)registered firms. There are two types of workers, low and high skill, and formal and informal firms have different skill intensities. 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. Informal firms are able to avoid all taxes and regulations, but they face an expected cost of being caught that is increasing in firm's size. Size and productivity are one-to-one and therefore more productive firms (in expectation) self-select into the formal sector and less productive firms enter the informal sector. The two margins of informality thus introduce a size-dependent distortion in the economy, which is shaped by existing regulations and government's enforcement technology (or lack thereof).[5] This structure is able to rationalize two prominent features of firm size distribution in developing countries: the predominance of very small firms, even in the formal sector; and the absence of a "missing middle" or other meaningful discontinuities (Hsieh and Olken 2014).

The proposed model encompasses the three leading views of informal firms (La Porta and Shleifer 2008, 2014), and is able to integrate them in a unified setting. The first view argues that the informal sector is a reservoir of potentially

[3] A separate literature stream studies the related issues of tax evasion and tax enforcement (e.g., the seminal work of Allingham and Sandmo 1972). This literature, however, focuses on tax systems and tax enforcement in developed countries (Slemrod and Yitzhaki 2002 provide a comprehensive review; for more recent studies, see Kleven et al. 2011 and Kuehn 2014).

[4] At least 40 percent of informal employment is located in formal firms in Brazil (Section I), and 44 percent in Mexico (de la Parra 2016). Perry et al. (2007) and Bertrand, Hsieh, and Tsivanidis (2015) provide evidence that the intensive margin is also relevant for other Latin American countries and India, respectively.

[5] In related work, Guner, Ventura, and Xu (2008) use a span-of-control framework to introduce a general size-dependent policy that increases the cost of capital. Garicano, Lelarge, and van Reenen (2016) analyze the French labor regulation, which increases labor costs discontinuously for firms with 50 employees or more.

productive entrepreneurs who are kept out of formality by high regulatory costs, most notably entry regulation. The second 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.[6] The third argues that informality is a survival strategy for low-skill individuals, who are too unproductive to ever become formal. Even though these views are seen as competing frameworks, I show that in fact they are not. They simply reflect heterogeneous firms choosing whether to comply with the relevant laws and regulations given the institutional framework they face.

I estimate the model using a simulated minimum distance estimator and matched employer-employee data on formal and informal firms in Brazil. I then use the estimated model to infer the relative size of each view in the data. The results show that the first view corresponds to 9.3 percent of all informal firms, while the second (the "parasite view") corresponds to 41.9 percent. The remaining firms correspond to low-skill entrepreneurs who are too unproductive to ever become formal and use informality as a survival strategy. These results therefore suggest that informal firms are to a large extent "parasite firms" and therefore eradicating them (e.g., through tighter enforcement) could in principle produce positive effects on the economy (e.g., Levy 2008). On the contrary, 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 positive impacts on informal firms that decide to formalize in the counterfactual equilibrium: an average increase of 13.3 percent in their net expected value at baseline. The effects are quite heterogeneous, ranging from 6.4 to 26.7 percent going from the first to the last quartile in the distribution of firm-level effects. For high productivity formal firms and all informal firms, this policy has negative impacts due to general equilibrium effects: greater entry increases competition and therefore equilibrium wages increase (mostly for high-skill workers), which hurts incumbents in both sectors. Increasing the costs of the extensive margin of informality benefits formal incumbents, in particular the low-productivity ones. This indicates that they are the most directly affected by the competition coming from informal firms. Increasing the costs of the intensive margin of informality is most harmful to low productivity formal firms, as they hire a large fraction of informal workers. Thus, these firms experience a substantial increase in their de facto labor cost.

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 essentially zero. Albeit puzzling at first, this result illustrates the importance of

[6] 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.

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 enforcement on the intensive margin increases: it generates a small reduction in the share of informal workers for both skill levels, but *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. These subtler policy impacts can only be uncovered if one explicitly considers the intensive margin, otherwise lower firm informality necessarily implies lower labor informality (and vice versa). As these results show, however, firm and labor informality can move in opposite directions.

Reducing entry costs also substantially increases the mass of active firms in the economy and leads to greater competition, output, and wages. The wage increase is concentrated on high-skill workers, which causes the skill premium in the economy to increase by 5 percentage points. This policy also generates a negative effect on aggregate TFP because of a negative composition effect, as it increases the presence of low-productivity firms. In contrast, increasing enforcement on the extensive margin nearly eradicates informal firms, which generates a large positive effect on aggregate TFP also due to composition effects, as this policy eliminates many small and unproductive informal firms. This positive effect on aggregate TFP more than compensates the reduction in the mass of active firms, and total output increases by 3.2 percent. Nevertheless, this policy generates an overall welfare loss in the economy. Therefore, lower informality can be, but is not necessarily associated to welfare gains.

The firm-level results are related to a literature stream that uses microdata to analyze the impact of different formalization policies in developing countries, among others: Monteiro and Assunção (2012) and Fajnzylber, Maloney, and Montes-Rojas (2011), who analyze tax reduction and simplification; Bruhn (2011), Kaplan, Piedra, and Seira (2011), and De Mel, McKenzie, and Woodruff (2013), who analyze the effects of reducing formal sector's bureaucratic entry costs; Rocha, Ulyssea, and Rachter (2018), who separately estimate the impacts of reducing entry costs and taxes; Almeida and Carneiro (2009, 2012) and De Andrade, Bruhn, and McKenzie (2014), who analyze the impacts of greater government auditing. The present approach, however, allows me to compute the full distribution of firm-level impacts 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 toward informality, which include Ulyssea (2010); Prado (2011); Charlot, Malherbet, and Terra (2015); D'Erasmo and Boedo (2012); and Leal Ordóñez (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, Narita, and Robin (2015), 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 and focuses on the analysis of labor markets for low-skill individuals. Their approach can thus be seen as complementary to the one proposed

in this paper, which focuses on firms' decisions, includes worker heterogeneity, and has the intensive margin of informality as its main innovation.

The next section presents the data and some key stylized facts. Section II presents the model, while Section III the estimation method and results. Section IV presents the counterfactual analysis and Section V concludes.

I. Informality Facts

A. *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)).[7] These definitions are used in the theory as well as in the data.

I use three datasets to conduct the empirical analysis. The two main ones 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 dataset 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'.[8] The ECINF is designed to be representative at the national level for firms with at most five employees.[9] To keep consistency across datasets, I only use data from 2003 for all of them, which is the last year available for the ECINF.

Although ECINF's sample size cap is not likely to be a problem when analyzing informal firms, which are predominantly small scale enterprises, it certainly is a binding restriction for the analysis of formal firms. I therefore use the *Registro Anual de Informações Sociais* (RAIS), an administrative dataset collected by the Ministry of Labor, which contains the universe of formal firms and workers. Besides providing complete information on formal firms and workers, the RAIS dataset is also useful to assess the quality of the ECINF (which is a survey). As Table 1 shows, both the size distribution and the composition across industries is remarkably similar in RAIS (which I restrict to firms with up to five employees, for comparability) and

[7] To register a firm in Brazil is a lengthy and costly process (online Appendix Table A.1). Besides these fixed registration costs, being a formal firms also implies ongoing costs such as taxes and red tape associated to tax payments, as well as other variable costs associated to the labor regulation.

[8] These are self-reported variables and naturally raise measurement error concerns. Nonetheless, IBGE has a long tradition in accurately measuring labor informality, and it has very strict confidentiality clauses, so the information cannot be used for auditing purposes. These features, associated to the high levels of informality observed in the data, increase the confidence that respondents are not systematically underreporting their informality status.

[9] 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 dataset.

Table 1—Comparing ECINF and RAIS

	RAIS (size $\leq$ 5)	Formal: ECINF	Informal: ECINF
Sector composition (*percent*)			
Services	41.9	42.4	50.5
Manufacturing	9.3	8.4	13.7
Retail	48.7	49.1	35.8
Size distribution (*number of workers*)			
Pc. 25	1	1	1
Pc. 50	2	2	1
Pc. 75	3	3	1
Pc. 95	5	5	3
Mean	2.2	2.4	1.3
Observations	1,310,166	5,257	30,627

Source: Author's own tabulations from RAIS and ECINF (2003)

ECINF, which is reassuring of ECINF's quality (online Appendix B contains the details of the construction of the datasets used).

Finally, I also use the National Household survey (PNAD), a repeated cross section that is representative at the national level, to compute statistics about formal and informal workers as well as aggregate labor market statistics (such as the share of informal workers). Relevant to this paper, the PNAD contains information on individuals' labor market outcomes (e.g., wages and employment), including formality status and detailed sociodemographic characteristics. It thus allows to compute wages controlling for observed heterogeneity, as well as to estimate formal-informal wage gaps controlling for workers' observable characteristics.

B. *Firms*

There are some well-established facts in the literature about informal firms in different countries: on average they have less educated entrepreneurs, are smaller both in terms of employees and revenues, pay lower wages, and earn lower profits relative to formal firms (e.g., La Porta and Shleifer 2008, 2014). These facts are also present in the Brazilian data (e.g., de Paula and Scheinkman 2011). The 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, online Appendix Figure C.1 provides evidence that they coexist even within narrowly defined industries (at the seven-digit level), which contradicts the notion that formal and informal firms operate in completely different markets.[10]

Using surveys for different countries, La Porta and Shleifer (2008, 2014) show 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

[10] Albeit using a different approach, Maloney (2004) argues that the formal and informal sectors are highly integrated in different Latin American countries.

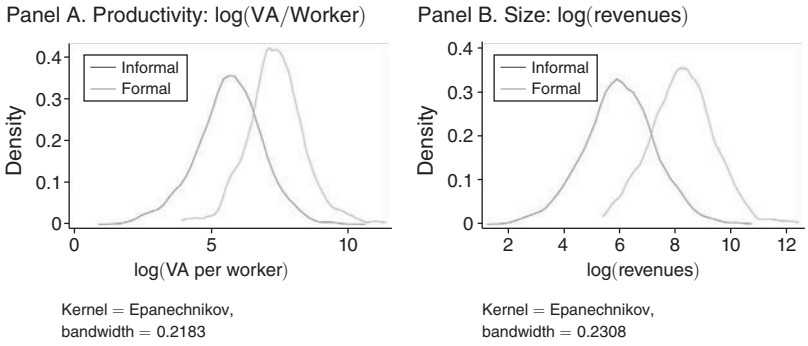

Figure 1. Productivity and Size Distributions among Entrants

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.

(value-added per worker) and size (log-revenues) for formal and informal firms at most one year old to proxy for entrants.[11] 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 between formal and informal firm size and productivity distributions. 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.[12]

C. *Workers*

A well-known stylized fact in the literature is that informal workers are on average less educated and less skilled than their formal counterparts. Accordingly, the share of informal workers is decreasing in workers' schooling level (e.g., Gasparini and Tornarolli 2009; Perry et al. 2007). Online Appendix Table C.1 shows the main descriptive statistics for workers using the PNAD, which basically confirm the same facts for Brazil. A particularly important regularity is that the share of informal workers is much higher among low-skill workers. Throughout the paper, I define as high-skill workers those who have at least completed high school and low skill as those with less than completed high school.

It is also a well-known fact that even after controlling for a myriad of observable characteristics (including schooling), there remains a substantial wage gap between formal and informal workers (e.g., Ulyssea 2010; Perry et al. 2007, and the references therein). Table 2 revisits this fact. In column 1, I use the National Household Survey (PNAD) to estimate the formal-informal wage gap controlling for workers'

[11] To obtain cleaner measures, I regress the log of value-added per worker 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.

[12] Meghir, Narita, and Robin (2015) show that this overlapping region is also present if one considers all formal and informal firms (and not only the entrants). The same is true for other countries that have comparable data, such as Mexico (Busso, Fazio, and Levy 2012) and other Latin American countries (Perry et al. 2007).

Table 2—Formal-Informal Wage Gaps Using Household Surveys and Matched Employer-Employee Data

	log(*wage*)		
	PNAD (1)	ECINF (2)	ECINF (3)
Formal contract (dummy)	0.2864	0.2413	0.0311
	(0.007)	(0.030)	(0.080)
High skill (dummy)	0.4583	0.1373	0.0921
	(0.006)	(0.031)	(0.0519)
Male (dummy)	0.2980	0.1256	0.1793
	(0.007)	(0.035)	(0.0434)
Age (years)	0.0740	0.0674	0.0365
	(0.002)	(0.007)	(0.010)
Age squared	−0.0008	−0.0007	−0.0003
	(0.000)	(0.000)	(0.000)
Observations	60,899	4,502	2,675
R^2	0.446	0.401	0.872
Firm fixed effects	No	No	Yes

Notes: PNAD is the National Household Survey, and ECINF is the matched employer-employee data for formal and informal firms and their employees. Variable *Formal* is a dummy for formal employee; *Skilled* is a dummy for workers with at least high school degree. All regressions control for five-digit industry classification. Robust standard errors in parentheses.

skill, gender, seven-digit industry dummies, age, and age squared. As the table shows, even after controlling for these observable characteristics, the formal-informal wage gap remains high (28.6 percent). Column 2 reproduces the same regression but now using data from the ECINF and the estimated wage gap is very close (which reinforces ECINF's quality). Finally, the third column exploits ECINF's matched employer-employee data to include firm fixed effects, thus estimating a *within-firm* formal-informal wage gap. As the table shows, the average wage gap between formal and informal workers disappears, becoming very small in magnitude and statistically nonsignificant.[13] If there is positive assortative matching in the economy, firm fixed effects will also capture workers' unobserved ability. Therefore, this result is expected if the following holds: (i) self-selection is the main driver of the wage gap between observably equivalent formal and informal workers; and (ii) these workers indeed perform the same tasks within the firm (conditional on skill).

D. *The Intensive and Extensive Margins of Informality*

Different papers have empirically examined the extensive margin of informality, showing that the probability of being informal strongly decreases with firms' size, usually measured by the number of employees (e.g., Perry et al. 2007). As panel A of Figure 2 shows, the same pattern is observed in the Brazilian data (see also de Paula and Scheinkman 2011). One possible rationale behind this fact is that larger

[13] The within-firm wage gap is identified from formal firms that hire both formal and informal workers, which explains the drop in sample size from the second to the third column. However, the fact that the wage gap disappears is not a consequence of the changes in the sample used. If one runs the same regression as in column 2 using column 3's sample, the wage gap remains positive and statistically significant, albeit smaller than the one estimated with the full sample.

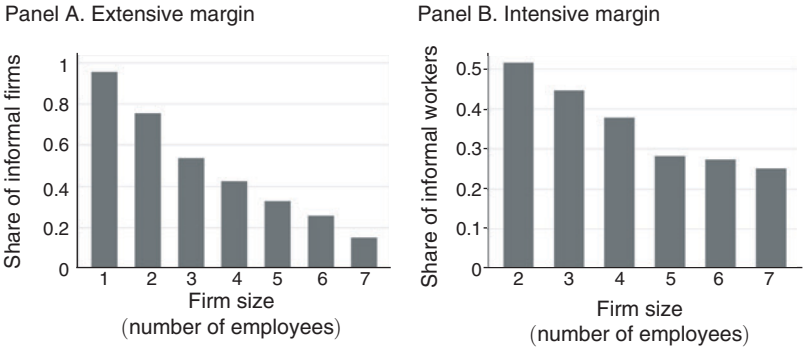

FIGURE 2. INFORMALITY MARGINS AND FIRMS' SIZE

Notes: Panel A shows the share of informal firms among firms with size $n = 1, \ldots, 7$ (where size is measured as number of employees). Panel B shows the average share of informal workers within formal firms, among firms with size $n = 2, \ldots, 7$.

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, panel B of Figure 2 shows that the intensive margin of informality is decreasing in firm's size, which is also true in other Latin American countries (Perry et al. 2007).

As for the empirical relevance of the intensive margin, the very few existing studies point to the sheer magnitude of this dimension of labor informality. In Mexico, for example, 44 percent of all informal employees are employed in formal firms, and they correspond to 23.4 percent of all workers employed in formal firms (de la Parra 2016). In India, Bertrand, Hsieh, and Tsivanidis (2015) show that large formal firms have increasingly used contract labor as a way to bypass the costs of labor regulation in India. From the firms' perspective, contract workers are analogous to informal workers within a formal firm. This form of labor relation corresponds to 36 percent of total employment among Indian establishments with more than 100 workers (Bertrand, Hsieh, and Tsivanidis 2015).

In the ECINF data, around 40 percent of informal employment is located in formal firms. Since the ECINF does not cover a large fraction of formal firms (due to its size cap of five employees), if anything this share is an underestimation of the importance of the intensive margin of informality in Brazil. Another way to assess the same issue is to examine the distribution of informal workers across firm sizes. The Brazilian Monthly Employment Survey (PME), which is a rotating panel that covers the six main metropolitan areas in Brazil, has categorized information on the size of workers' firms, as well as workers' formality status. Online Appendix Table C.2 uses data from the PME to show that 52 percent of all informal workers are employed in firms with 11 employees or more (Perry et al. 2007 show similar evidence for other Latin American countries). 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 reinforce that there is a large fraction of informal workers who are employed in formal firms.

II. Theory

Motivated by the facts previously discussed, this section develops an equilibrium model where firms can exploit both the extensive and intensive margins of informality. The model is similar in spirit to the seminal contribution of Melitz (2003) but with some important changes in the entry and production structures (to be discussed ahead), as well as the addition of the two 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.[14] I start with a simpler version of the model where workers are homogeneous. I derive the main results using this simpler model and then extend it to include two types of workers, low and high skill, following the same definition used in the previous section. As I discuss below, the main insights of the model with homogeneous workers are carried over to the model with two skill levels. In both versions of the model, formal and informal employees perform the exact same tasks within the firm (conditional on their skill level, when workers are heterogeneous). Thus, there is no wage difference between formal and informal workers, conditional on their skill level.[15]

A. *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 $\tau_i(\ell)$, which 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., de Paula and Scheinkman 2011).[16] Informal firms' profit function is thus given by

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

where the price of the final good is normalized to 1.

Formal incumbents must comply with taxes and regulations, but they can hire informal workers to avoid the costs implied by the labor legislation.[17] The hiring

[14] As argued in Section I, 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.

[15] The model therefore abstracts from all non-wage benefits included in a formal contract, such as unemployment insurance (for a more detailed treatment of formal workers' total compensation, see Meghir, Narita, and Robin 2015).

[16] In the online Appendix D.1 I show that the formulation with the general cost function, $\tau_i(\cdot)$, can be specialized to a formulation that explicitly accounts for a detection probability.

[17] It is worth noting that concavity in the production function plays an important role in rationalizing the existence of the intensive margin of informality. In the presence of constant returns, firms could in principle divide their operations into formal and informal establishments, which would only hire formal and informal workers, respectively.

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 $\tau_{fi}(\cdot)$, $\tau'_{fi}, \tau''_{fi} > 0$. 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 payroll 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).[18] Formal firms' profit function can be written as follows:

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

and

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

where τ_y denotes the revenue tax. Incumbents in both sectors must pay a per-period, fixed cost of operation, which is denoted by $\overline{c}_s$, $s = i,f$. 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 $\pi_s(\theta, w) = \Pi_s(\theta, w) - \overline{c}_s$.

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 highly tractable formulation is able to capture the main facts discussed in the previous section regarding both margins of informality.[19]

B. *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. Assume that G is

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

[19] 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. One possible way around this limitation would be to assume that firms above a given size threshold are inspected with probability 1. However, this would add yet another parameter to estimate and would not change the main implications of the model, as it already captures well the behavior of the *share* of informal workers within formal firms.

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: $E_f > E_i$.[20]

After entry occurs, firms draw their actual productivity from the conditional cumulative distribution function $F(\theta|\nu)$, which is the same in both sectors and independent across firms. The function $F(\theta|\nu)$ 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. Importantly, once firms draw their productivity θ, it remains constant forever and firms face an exogenous exit probability denoted by κ_s, $s = i,f$. Thus, similarly to Melitz (2003), there is endogenous entry but exogenous exit, as there is no actual dynamics after entry occurs.[21] However, the entry structure has a fundamental difference, as firms are ex ante heterogeneous and only realize their actual productivity after entry occurs. Therefore, the model allows for the possibility of overlap between formal and informal productivity distributions, which is an important regularity in the data.[22] By contrast, the fully static models without uncertainty imply perfect sorting and no overlap between formal and informal firms' productivity and size distributions, which is at odds with the data (as shown in Section I).

If firms are surprised with a low productivity draw $\theta < \overline{\theta}$, where $\pi_s(\overline{\theta}, w) = 0$, they decide to exit immediately without producing. Aggregate prices remain constant in steady-state equilibria and since firms' productivity also remains 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 1. Note, however, that the exit probability κ_s could also be interpreted as a sector-specific discount rate, which could reflect, for example, differential borrowing rates.

The expected value of entry for a firm with pre-entry signal ν is thus given by

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

[20] The difference between E_f and E_i can be interpreted as the costs implied by the regulation of entry into the formal sector. The latter includes direct costs, such as red tape and fees, but can also be interpreted as the monetization of expected costs, such as firing costs. 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.

[21] This structure also implies that initial formalization decision is permanent and there is no transition between informal and formal status. La Porta and Shleifer (2008) show that, in a sample of 14 Latin American countries, on average 91.2 percent of firms registered upon formation, which suggests that there is limited firm transition between the informal and formal sectors.

[22] Even though this entry structure provides a reasonable rationalization for the overlap in productivity distributions among formal and informal *entrants*, it should be seen as a reduced-form approximation to the mechanisms that might produce this overlap among older firms.

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 $V_i^e(\nu, w) - E_i > \max\{V_f^e(\nu, w) - E_f, 0\}$. If entry in both sectors is positive, the following entry-conditions hold:

$$V_i^e(\overline{\nu}_i, w) = E_i,$$

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

where $\overline{\nu}_s$ is the pre-entry productivity of the last firm to enter sector $s = i, f$. Online Appendix D.2 shows that the effective, post-entry productivity distributions in both sectors can be derived as functions of these thresholds.

C. *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. Consumers do not derive any disutility from work and cannot save, so they simply consume all of their income. Total consumption constitutes the natural welfare measure in this context, which is given by $w\overline{L} + \Pi + T$. The Π denotes total profits in the economy net of total entry costs, $M_f E_f + M_i E_i$, where $M_i = \left[G(\overline{\nu}_f) - G(\overline{\nu}_i)\right] M$ and $M_f = \left[1 - G(\overline{\nu}_f)\right] M$ are the measures of entrants into the informal and formal sectors, respectively. The $w\overline{L}$ denotes total wages ($\overline{L}$ is the labor endowment), and T denotes tax revenues, which are directly transferred to the household.[23]

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

$$\mu_s = \frac{1 - F_{\theta_s}(\overline{\theta}_s)}{\kappa_s} M_s \,, \tag{5}$$

where μ_s denotes the mass of active firms in sector s. In words, condition (5) 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) labor market clears, $L_i + L_f = \overline{L}$; (ii) The zero profit cutoff (ZPC) condition holds in both sectors, $\theta \geq \overline{\theta}_s$ where $\pi_s(\overline{\theta}_s, w) = 0$; (iii) the free entry condition holds in both sectors, with equality if $M_s > 0$; and (iv) both sectors' size remains constant (expression (5)). Online Appendix D.3 shows that the equilibrium exists and it is unique.

[23] Note that total profits also exclude the costs of informality, $\tau_s(\cdot)$, $s = i, f$. These are assumed to be wasted resources as a consequence of laws and regulations that are imperfectly enforced.

D. *Adding Worker Heterogeneity*

This section extends the basic model presented in the previous sections to include two types of workers, low and high skill. The basic set up remains the same but now the labor input is a constant elasticity of substitution (CES) aggregation of two types of workers:

$$\ell_s \;=\; \left(\eta_s l_1^{\rho} + \left(1-\eta_s\right) l_2^{\rho}\right)^{\frac{1}{\rho}},$$

where l_1 denotes high-skill workers and l_2 low-skill workers; $s = i,f$ indexes sectors; the η_s denotes the share parameter in each sector; and the ρ is a common elasticity of substitution parameter. As in the previous sections, firms' output is given by $y(\theta,\ell_s) = \theta q(\ell_s)$.

The cost of the extensive margin of informality continues to take the form of a labor distortion denoted by $\tau_i(\ell_i)$, which is increasing in firm's composite employment ℓ_i. Informal firms' profit function is given by

$$\Pi_i(\theta, w_1, w_2) \;=\; \max_{l_1, l_2}\left\{\theta q(\ell_i) - \tau_i(\ell_i)\left(w_1 l_1 + w_2 l_2\right)\right\}, \tag{6}$$

where the price of the final good is normalized to 1.

Formal firms also face an increasing and convex expected cost to hire informal workers, which can differ across workers' skill levels: $\tau_{fs}(l_s)$, $\tau'_{fs}, \tau''_{fs} > 0$, where $s = 1,2$. The rationale for this specification is to account for the fact that formal firms may face different costs and benefits to formalize low- and high-skill workers, which are captured by different cost functions.[24] The same is not relevant for informal firms as they keep their entire business at the margin of all relevant laws and regulations.

Since the cost functions differ across skill levels, formal firms will have different thresholds to start hiring low- and high-skill formal workers, which are denoted by $\tilde{l}_s$.[25] If the labor quantity that maximizes formal firm's profit is such that $l_s^* \leq \tilde{l}_s$, then it will only hire informal workers of skill level s. If $l_s^* > \tilde{l}_s$, the firm hires $\tilde{l}_s$ informal workers and $l_s^* - \tilde{l}_s$ formal workers. The functions $\tau_{fs}(\cdot)$ are parameterized and estimated, so the data will determine whether the thresholds $\tilde{l}_1$ and $\tilde{l}_2$ are different. If they are the same, then the formal firm will either hire all of its labor force informally, or will hire some fraction of both types of workers formally. The data shows that the share of informal workers is higher among low-skill workers, so it is likely that $\tilde{l}_1 < \tilde{l}_2$. If this is the case, formal firms can be in one of three possible situations: (i) hire all of its workers informally, if $l_s^* \leq \tilde{l}_s$, $s = 1,2$; (ii) hire all of its low-skill workers informally but some high-skill workers formally, if

[24] For example, workers of different skill levels can have different probabilities of denouncing firms to the labor authorities, which would imply different expected costs of hiring informally.

[25] The reasoning is analogous to the single factor model. The marginal cost of hiring informal workers $\tau'_{fs}(\cdot)w_s$ is strictly increasing, while the marginal cost of hiring formal workers $(1+\tau_w)w_s$ is constant. Hence, there is an unique value of $\tilde{l}_s$ such that $\tau'_{fs}(\tilde{l}_s) = 1+\tau_w$.

$l_1^* > \tilde{l}_1$, $l_2^* \leq \tilde{l}_2$; and (iii) hire some formal workers of both skill levels, if $l_s^* > \tilde{l}_s$, $s = 1, 2$. The profit maximization can thus be written as follows:

$$\Pi_f(\theta, \mathbf{w}) = \max_{l_1, l_2} \left\{ (1 - \tau_y)\, \theta\, \ell^{\alpha} - C(l_1, l_2) \right\} \tag{7}$$

and

$$C(l_1, l_2) = \begin{cases} \tau_{f1}(l_1)\, w_1 + \tau_{f2}(l_2)\, w_2 & \text{for } l_s \leq \tilde{l}_s, s = 1, 2 \\ \tau_{f1}(\tilde{l}_1)\, w_1 + (1 + \tau_w)\, w_1 (l_1 - \tilde{l}_1) + \tau_{f2}(l_2)\, w_2 & \text{for } l_1 > \tilde{l}_1, l_2 \leq \tilde{l}_2 \\ \sum_{s=1,2} \left\{ \tau_{fs}(\tilde{l}_s)\, w_s + (1 + \tau_w)\, w_s (l_s - \tilde{l}_s) \right\} & \text{for } l_s > \tilde{l}_s, s = 1, 2. \end{cases}$$

Online Appendix E discusses the solution to formal firms' problem in each of these cases. It is worth highlighting that despite the substantial changes implied by the inclusion of two levels of skills, this extension does not alter the nature of firms' problems, as the properties of the value functions are not altered. Thus, the decisions to enter either sector and whether to stay active after entry occurs remain the same, and so are the equilibrium conditions. The only difference is that now labor market clearing involves two equations, as labor supply and demand must equate for both skill levels.

III. Estimation

This section discusses the estimation of the full model with worker heterogeneity presented in Section II. The model describes firms' decisions regarding entry, production, and compliance with regulations in an equilibrium setting. To perform counterfactual analysis of policy changes, it is necessary to estimate all objects in the model's structure. I estimate the model using a two-step Simulated Minimum Distance (SMD) estimator. This approach combines direct estimation and calibration from micro- and macro-data in the first step, with the SMD estimator itself in the second step.

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.[26] The next section describes the parameterization used, while Section IIIB describes the estimation method, as well as discusses identification and the model's fit.

A. *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, $\tau_i(\cdot)$ and $\tau_{f,k}(\cdot)$, were left unspecified. This section completes the model's parameterization by

[26] 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 (e.g., Heckman 2001; Ichimura and Taber 2002). However, ex ante policy evaluations typically require the full specification of a behavioral model (e.g., Keane, Todd, and Wolpin 2011).

assuming specific functional forms for these objects. Starting with the pre-entry productivity distribution, it is assumed to be Pareto:

$$F_\nu(\nu \geq x) = \begin{cases} \left(\frac{\nu_0}{x}\right)^\xi & \text{for } x \geq \nu_0 \\ 1 & \text{for } x < \nu_0. \end{cases} \tag{8}$$

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. The product of a log-normal and a Pareto random variable produces a Pareto-Lognormal distribution, which was first introduced by Colombi (1990) and has been increasingly used in different applications (e.g., Rothschild and Scheuer 2016). This is a three-parameter distribution that has a log-normal body and a Pareto right tail, which fits well many salient features of firm size distribution (e.g., Luttmer 2007).

As for production, I assume the span-of-control formulation: $y(\theta, \ell_s) = \theta \ell_s^\alpha$, where $\alpha < 1$ and ℓ_s is the CES aggregation of low and high-skill labor in sector $s = i,f$. The cost functions of both margins take a very simple functional form: $\tau_i(\ell_i) = \left(1 + \frac{\ell_i}{b_i}\right)$, where $b_i > 0$ and $\tau_{f,k}(l_k) = \left(1 + \frac{l_k}{b_{fk}}\right) l_k$ and $k = 1, 2$ denotes workers' skill level. Finally, I assume that the per-period, fixed costs of operation are a function of the equilibrium wage for unskilled workers, which makes the exit margin more meaningful since it now responds to market conditions.[27] The fixed costs are determined as follows: $\overline{c}_s = \gamma_s w_2, 0 < \gamma_s \leq 1$.

I partition the vector of parameters into two subvectors, $\Gamma = \{\psi, \varphi\}$ and proceed in two steps. In the first step, I determine the following parameters: $\psi = \{\tau_w, \tau_y, \kappa_f, \nu_0, \gamma_f\}$. The tax rates are set to their statutory values: $\tau_w = 0.375$ and $\tau_y = 0.293$. The value of τ_w corresponds to the main payroll taxes, namely, employer's social security contribution (20 percent), direct payroll tax (9 percent), and severance contributions (FGTS), 8.5 percent. The value of τ_y includes two VAT-like taxes: the IPI (20 percent) and PIS/COFINS (9.25 percent), which correspond to the federal taxes only. These statutory values can be easily obtained in the compilation by the World Bank's Doing Business initiative. Since state level value-added taxes vary substantially across states and there is a cumbersome system of tax substitution across the production chain, I exclude those from the parameterization.[28] The exit probability in the formal sector is $\kappa_f = 0.129$, which is estimated using the panel structure in the RAIS dataset. 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'

[27] In a fully dynamic model with endogenous exit, the equilibrium wage affects firms' survival probability. Even though this channel is absent in the present model, this specification allows wages to affect firms' survival in equilibrium.

[28] These taxes can be sizable in some states, in which case this assumption would imply that the parameterization underestimates the overall tax burden. On the other hand, the model does not have intermediate inputs and therefore the effective value-added tax is lower for some firms, which would imply that the tax burden is overestimated. It is unclear a priori which effect dominates.

minimum size is one employee, while formal sector's fixed cost of operation ($\gamma_f = 0.5$) is set to be one-half of themonthly wage.

B. *Estimation Method*

I take the parameters defined in the first step as given and proceed to use a Simulated Minimum Distance (SMD) estimator to obtain the reminder parameters of the model. There are 13 parameters to be estimated in this second step:

$$\varphi = \left\{\kappa_i, \gamma_i, b_i, b_{f,1}, b_{f,2}, \eta_f, \eta_i, \rho, \xi, \alpha, E_f, E_i, \sigma\right\}.$$

For a given parameter vector (Γ), wages (w_1 and w_2) and individual productivity shocks (ν_j and ε_j), one can use the model to completely characterize firms' behavior. The estimator proceeds by using the model to generate simulated datasets of formal and informal firms and computing the 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.[29]

Let $\hat{m}_N = \frac{1}{N}\sum_{i=1}^{N} m_i$ denote the vector of moments computed from data, which can include, for example, the share of informal workers among low-skill workers. Let the simulated counterparts of these moments to be denoted by $\tilde{m}_S(\varphi;\psi) = \sum_{s=1}^{S} \tilde{m}_s(\varphi;\psi)$, which are constructed based on a set of S series of simulated data. Define $g_{NS}(\varphi;\psi) = \hat{m}_N - \tilde{m}_S(\varphi;\psi)$; the estimator is then given by

$$\hat{\varphi} = \arg\min_{\varphi} Q(\varphi;\psi) = \left\{g_{NS}(\varphi;\psi)' \hat{\mathbf{W}}_N g_{NS}(\varphi;\psi)\right\}, \tag{9}$$

where $\hat{\mathbf{W}}_N$ is a positive, semi-definite $r \times r$ matrix, where r is the length of the vector of moments $\hat{m}_N$. Under the suitable regularity conditions, the estimator is consistent and asymptotically normal and the weighting matrix $\hat{\mathbf{W}}_N$ is chosen optimally in order to minimize the asymptotic covariance.[30] Online Appendix F discusses the details of the estimation, including a brief description of the required regularity conditions, the asymptotic variance-covariance matrix, and the computation of $\hat{\mathbf{W}}$.

Moments and Identification.—I use 16 moments from the data to form the vector $\hat{m}_N$,[31] which are the following: (i) informality share among low-skill, high-skill, and all employees (data source: PNAD); (ii) overall share of informal firms and by firm size for 1–2, 3–4, and 5–10 employees (data sources: ECINF and

[29] The technical details are discussed in online Appendix F.2. The interested reader can find a systematized discussion in Gouriéroux and Monfort (1996) and Adda and Cooper (2003).

[30] Altonji and Segal (1996) argue that the Optimal Minimum Distance (OMD) estimator can be severely biased in small samples, but they show that the bias dissipates with sample size. Given that the sample sizes used here largely exceed those for which the bias is shown to be negligible, I use the OMD estimator due to its desirable asymptotic properties.

[31] As discussed in Section I, the datasets used are the following: ECINF, a repeated cross section of small firms; RAIS, an administrative dataset from the Ministry of Labor; and PNAD, the National Household Survey, a repeated cross section that is representative at the national level.

RAIS); (iii) average share of informal workers within formal firms with size 2–3 and 4–5 employees (data source: ECINF); (iv) the share informal firms with less than 2 and less than 5 employees (data source: ECINF); and (v) the share of formal firms with up to 5, 5 to 10, 11 to 20, 21 to 50, and more than 50 employees (data source: RAIS).

A crucial question is whether these moments are a good choice and if they allow me to identify the parameters of the model. In online Appendix Section F.3, I follow the analysis in Adda, Dustmann, and Stevens (2017) to address this question. The basic idea is that the objective function should not be flat in the region around the vector of estimated parameters. If it is, this would raise identification concerns, as would suggest that the moments chosen are not particularly informative about the model's parameters. As online Appendix Figure F.1 shows, the objective function (expression (9)) is sensitive to small changes in the vector of estimated parameters.[32]

Even though the goal of this section is not to provide a constructive argument for identification, in what follows I provide further intuition on how the variation in the data, combined with the model's structure, identify the parameters of the model. The shape parameter of the Pareto distribution, ξ, and the variance of the post-entry shock are determined by the moments of firm size distribution in the formal and informal sectors. The latter is also disciplined by the degree of overlap between formal and informal firm size distributions, as it will determine how much post-entry dispersion there is in productivity. Put differently, if there was no uncertainty, the model would imply that the two size distributions are disjoint. It is worth emphasizing that the importance of firm size distributions to identify the parameters associated to the productivity distribution comes from the fact that the model implies a one-to-one relationship between productivity and firms' size.[33]

The parameter that governs the cost of informality, b_i, is identified by the variation in the share of informal firms by firm size. Since the function $\tau_i(\ell_i)$ is increasing in firms' size, it implies that the share of informal firms will be decreasing as firms get larger simply because it becomes increasingly costly to do so. The intensity with which the cost of informality increases with firm size, and therefore the intensity with which the share of informal firms declines, crucially depends on b_i. Analogously, the share of informal workers in formal firms per firm size is central to identify the costs of formal firms hiring informal workers of both skill levels, $b_{f,k}$, $k = 1, 2$. Additionally, these parameters are disciplined by the overall share of informal workers by skill level in the economy. Since the share of informal workers is much larger among low-skilled workers, the model partially loads that into $b_{f,2}$ and tends to produce estimates such that $b_{f,2} > b_{f,1}$.

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. Importantly, the cost of entry into the formal sector E_f affects the left tail of the size

[32] It is worth noting that this analysis, by construction, relates to local and not global identification, as it focus on the region around the vector of estimated parameters.

[33] The exception to this correspondence are the informal firms that bunch around the transition threshold into the formal sector.

Table 3—Parameter Values

Parameter	Description	Source	Value	SE
First step				
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 step				
α	Cobb-Douglas coefficient	Estimated	0.605	0.008
b_{f1}	Intensive mg. cost: skilled	Estimated	2.61	0.702
b_{f2}	Intensive mg. cost: unskilled	Estimated	4.94	0.864
b_i	Extensive mg. cost	Estimated	5.01	0.301
κ_i	Informal sector's exit probability	Estimated	0.381	0.040
γ_i	Per-period fixed cost of operation (informal)	Estimated	0.248	0.065
ξ	Pareto's shape parameter	Estimated	3.08	0.073
σ	Post-entry shock variance	Estimated	0.245	0.006
ρ	CES elasticity parameter	Estimated	0.290	0.097
η_I	Informal CES share parameter	Estimated	0.481	0.026
η_F	Formal CES share parameter	Estimated	0.593	0.015
E_f[a]	Formal sector's entry cost	Estimated	4,286.2	502.1
E_i[a]	Informal sector's entry cost	Estimated	2,023.4	353.9

Note: Formal and informal sector's estimates and SE expressed in R$ of 2003.

distribution in the formal sector: if it is too high, there will be fewer small formal firms and the size distribution will be more shifted to the right relatively to informal sector's. These differences allow the model to separate the effects of E_f and κ_i on the degree of informality in the economy, as the latter affects all firms equally. Finally, given γ_f (determined in the first stage), γ_i is one of the determinants of post-entry survival in the informal sector and is directly connected to the importance of very small firms in the size distribution of informal firms.

Estimates and Model Fit.—Table 3 shows the values of all parameters. The estimates show that formal sector's entry cost is more than twice informal sector's. The 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. The parameter that governs the cost function of the intensive margin of informality for low-skill workers, b_{f2}, is nearly twice as large as the one for high-skill workers, b_{f1}. This indicates that it is easier to hire low-skill workers informally than it is to hire high-skill ones. This is consistent with the fact that in the data the share of informal workers is higher among low- than high-skill workers. Pareto's shape parameter, ξ, indicates that the pre-entry productivity distribution is skewed to the right.

Table 4 shows how the model performs compared to some of the targeted moments in the data. The model matches the share of informal firms and the share of informal workers well, for all workers as well as by skill level. However, it understates the share of informal firms with up to two employees. The same does not happen with the size distribution in the formal sector, which the model is able to replicate well. Indeed, Figure 3 shows that the model fits well the entire firm size distribution in

TABLE 4—MODEL FIT

Moments	Source	Model	Data
Share of informal workers			
All	PNAD	0.354	0.354
Low-skilled	PNAD	0.428	0.424
High-skilled	PNAD	0.269	0.260
Share of informal firms	ECINF + RAIS	0.687	0.698
Size distribution: informal firms			
≤2 employees	ECINF	0.772	0.957
≤5 employees	ECINF	0.996	0.998
Size distribution: formal firms			
≤5 employees	RAIS	0.704	0.701
6–10 employees	RAIS	0.146	0.141
11–20 employees	RAIS	0.081	0.083
21–50 employees	RAIS	0.046	0.048
50+	RAIS	0.024	0.027

Notes: The PNAD is the National Household Survey, ECINF (*Pesquisa de Economia Informal Urbana*) is a repeated cross section of small firms (up to five employees) and RAIS (*Relacao Anual de Informacoes Sociais*) is an administrative dataset collected by the Ministry of Labor. All moments are computed using data from 2003, which is the last year ECINF is available.

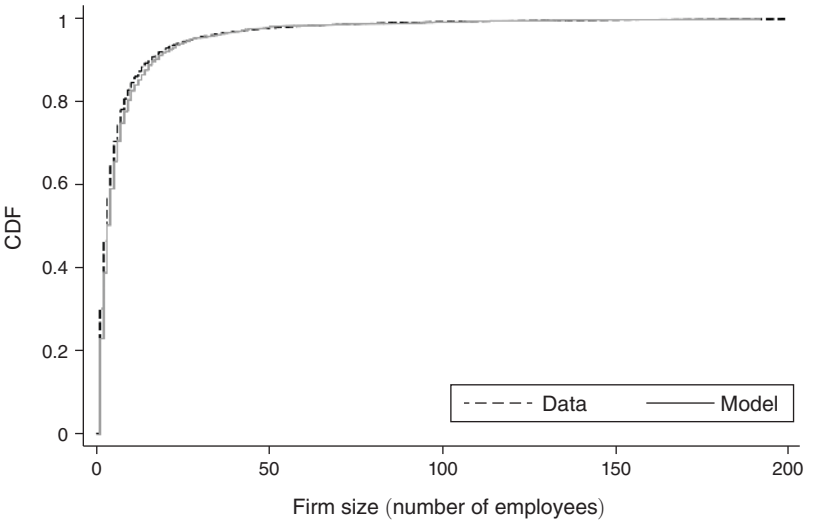

FIGURE 3. FIRM SIZE DISTRIBUTION IN THE FORMAL SECTOR: DATA VERSUS MODEL

Notes: The data source is a 25 percent random sample of the RAIS dataset. Firm size is measured as number of employees.

the formal sector, and not only the targeted moments.[34] Figure 4 shows the share of high-skill workers per firm size both in the data and the model, which are also moments not targeted in the estimation. The model is able to reproduce well the

[34] 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 online Appendix F.2).

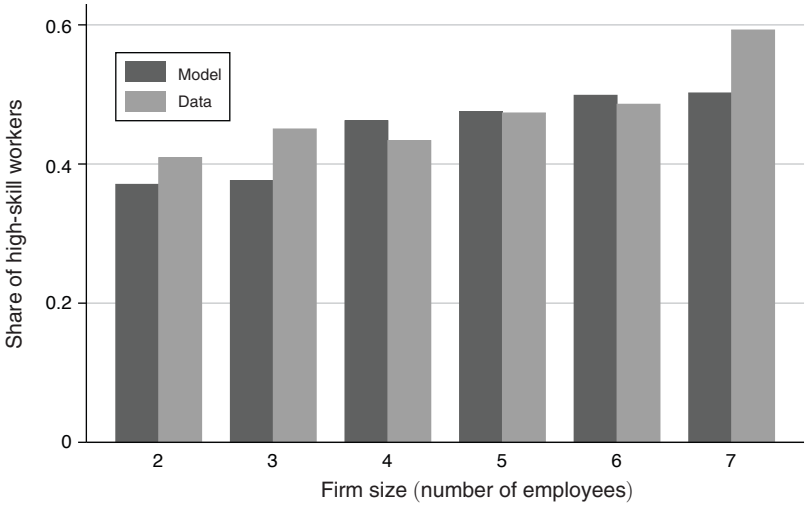

FIGURE 4. SHARE OF HIGH SKILL WORKERS PER FIRM SIZE: DATA VERSUS MODEL

Source: ECINF survey

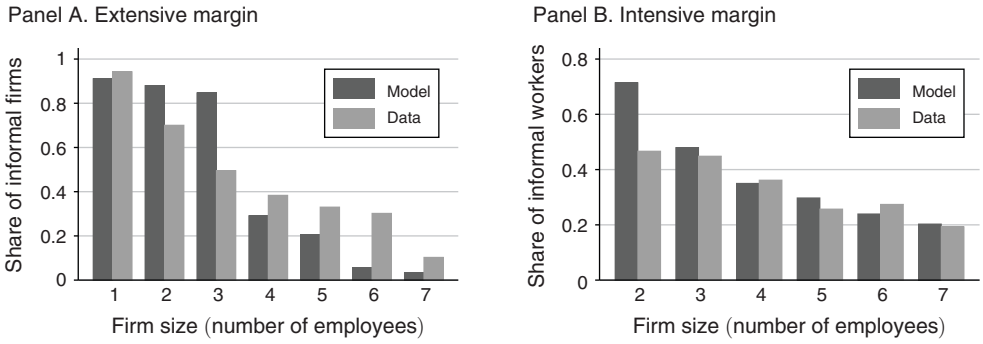

FIGURE 5. INFORMALITY MARGINS: DATA VERSUS MODEL

Notes: Panel A shows the share of informal firms among firms with size $n = 1, \ldots, 7$ (where size is measured as number of employees). Panel B shows the average share of informal workers within formal firms, among firms with size $n = 2, \ldots, 7$.

patterns found in the data, although it underestimates by around 9 p.p. this share among firms with seven employees.[35] Finally, Figure 5 shows the fit of the model regarding the behavior of the two informality margins, discussed in Section I. The model reproduces well the behavior observed in the data, in particular the intensive margin. It tends to overestimate the decline of the extensive margin as firms grow (panel A) but the predicted behavior is nevertheless similar to what is observed in the data.

[35] As for the share of high-skill workers in formal and informal firms, the model is also able to reproduce well the behavior observed in the data. The share of high-skill workers in formal firms is 53.4 percent in the data and 50.6 percent in the model; as for informal firms, 31.3 percent in the data and 35.3 percent in the model.

IV. Counterfactual Analysis

A. *The Distribution of Informal Firms' Types*

There exist three main views in the literature about the role of informal firms in economic development (see La Porta and Shleifer 2008, 2014, for a discussion). These views imply very different perceptions of how informal firms affect the economy and what are the best policies to tackle informality. Even though they are seen as competing frameworks for understanding informality, in this section I show that in fact they are not, they simply reflect firm heterogeneity in the informal sector. Hence, these views are complementary and not competing frameworks for understanding informality. The crucial question is therefore what is the relative importance of each view in the data.

To answer this question, I propose a simple taxonomy of informal firms based on these views and use the estimated model to back out the distribution of informal firm types. The starting point is to establish a precise definition of each type, which comes directly from these views. The first type corresponds to the *Survival view*, which refers to 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. The second is the *Parasite view* (type 2), which corresponds to informal firms that are productive enough to enter the formal sector once entry barriers are removed, but *choose* not to do so because it is more profitable to operate in the informal sector. Finally, the *De Soto's view* (type 3) corresponds to potentially productive informal firms that are kept out of formality by high entry costs. If these were removed, they would enter the formal sector 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 counterfactual 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 choose not to enter the formal sector. Type 2 firms are productive enough to survive in the formal sector (once entry barriers are removed), while Type 1 firms are simply not productive enough to enter the formal sector, even if regulatory entry costs are eliminated.

Given this reasoning, the relevant thought experiment is to ask how firms that decide to enter the informal sector in the baseline economy would respond to an intervention that equalizes entry costs in the formal and informal sectors ($E_f = E_i$). I use the estimated model to obtain, for each firm with pre-entry productivity ν, the baseline net expected value of entering the formal and

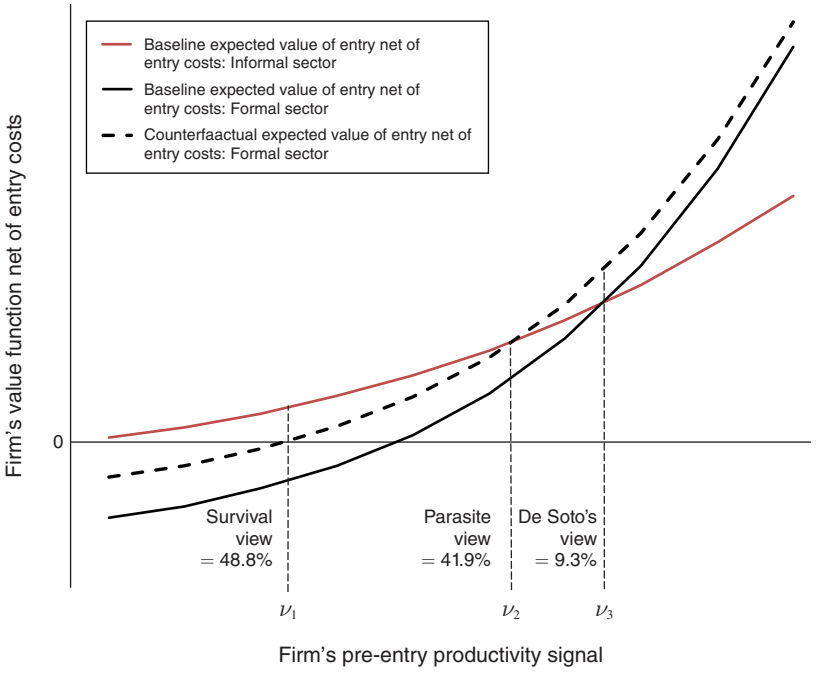

FIGURE 6. THE DISTRIBUTION OF INFORMAL FIRMS TYPES IN THE DATA

Notes: The figure shows, for each productivity level, firms' expected value function net of entry costs in the formal and informal sectors, $V_f^e(\nu) - E_f$ and $V_i^e(\nu) - E_i$, respectively. The third curve displays the net expected value of entering the formal sector in a scenario where its entry costs are equalized to informal sector's ($E_f = E_i$): $V_f^{e,c}(\nu) - E_i$.

informal sectors, respectively: $V_f^e(\nu) - E_f$ and $V_i^e(\nu) - E_i$. I then simulate the counterfactual scenario where entry costs into the formal sector are equalized to informal sector's ($E_f = E_i$), and compute for all firms the counterfactual expected value of entering the formal sector once entry costs are removed, which is given by $V_f^{e,c}(\nu) - E_i$, where the superscript c denotes the counterfactual scenario. Figure 6 displays the corresponding curves.

The baseline curves for the formal and informal sectors intersect each other at $\nu = \nu_3$, and all firms with $\nu \geq \nu_3$ will always choose to be formal, as their expected value of entry is higher than in the informal sector. Firms with pre-entry productivity $\nu \in [\nu_2, \nu_3)$ are the De Soto's (type 3) firms: in the counterfactual scenario where entry barriers into the formal sector are removed, these firms enter the formal sector, improve their performance, and achieve higher profits. Firms with pre-entry productivity $\nu \in [\nu_1, \nu_2)$ correspond to Parasite (type 2) firms. They are productive enough to enter the formal sector once entry costs are removed, as their expected value of entry 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 $\nu < \nu_1$ are the Survival (type 1) firms, which are not productive enough to enter the formal sector even when the fixed costs of formalization are removed.

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 two elements: (i) the underlying pre-entry productivity distribution F_ν and (ii) the determinants of

firms' sorting between sectors. As discussed in the previous section, the parameters that govern (i) 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 indicate 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 9.3 percent 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 noncompliance correspond to 41.9 percent of all informal firms. The remaining firms, 48.8 percent, 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.

B. *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) reducing the cost of entry into the formal sector to make it the same as in the informal sector; (ii) a 20 p.p. cut in the payroll tax, which corresponds to eliminating employer's social security contribution; (iii) increasing the cost of being an informal firm (the extensive margin); and (iv) increasing the cost of formal firms hiring informal workers for both skill levels (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 parameters b_i and $b_{f,k}$, $k = 1, 2$.[36] In what follows, I analyze the effects at the firm and aggregate levels separately.[37]

Impacts on Firms.—To analyze the effects on firms, I contrast firms' outcomes in the counterfactual and baseline scenarios. I define firm's expected value of entry net of entry costs as the outcome of interest, $V_s^e(\nu) - E_s$, where $s = f, i$. The firm-level treatment effect is then given by

$$\Delta(\nu) = \log\big(V_s^{e,c}(\nu) - E_s^c\big) - \log\big(V_s^{e,b}(\nu) - E_s^b\big), \tag{10}$$

where ν denotes firm's pre-entry productivity signal, the superscripts b and c denote the baseline and counterfactual scenarios, respectively, and $s = i, f$ indexes firm's sector (formal and informal). Note that one of the counterfactual scenarios involves changing formal sector's entry cost, and therefore entry costs are also indexed by b and c in the expression above.

In order to organize the results, I divide firms into three basic groups according to their choices: (i) "always formal," which are those that choose to be formal in the

[36] I consider an extreme case where the parameters b_i and $b_{f,k}$ are set close to zero.

[37] For each counterfactual exercise, I compute the new equilibrium assuming policy invariance of the remaining parameters, including formal and informal firms' exit rates. In a fully dynamic model, however, firms' exit decision would be endogenous and therefore could also respond to policy changes.

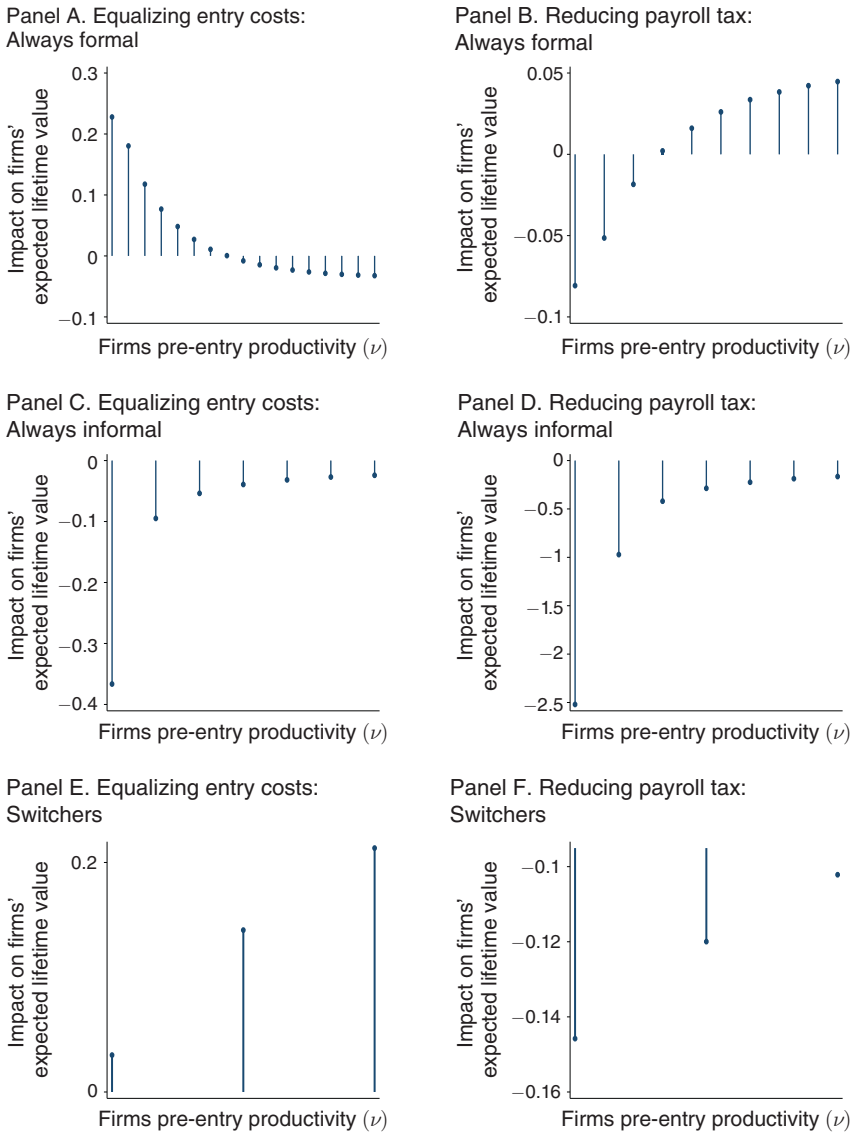

FIGURE 7. PROFILES OF FIRM-LEVEL EFFECTS: REDUCING REGULATORY COSTS

baseline and counterfactual scenarios; (ii) "always informal," which are those that choose to be informal in both scenarios; and (iii) "switchers," which are those that choose to enter the informal sector in the baseline scenario but choose to enter the formal sector in the counterfactual scenario. I start by analyzing the profiles of firm level effects across different productivity levels within each of these groups of firms and for the four formalization policies considered. For each policy and group, I compute the average effect at each point of firms' productivity grid. Figure 7 shows the profiles for the policies that reduce regulatory costs (entry costs and payroll tax), which correspond to policies (i) and (ii) above. Figure 8 shows the profiles

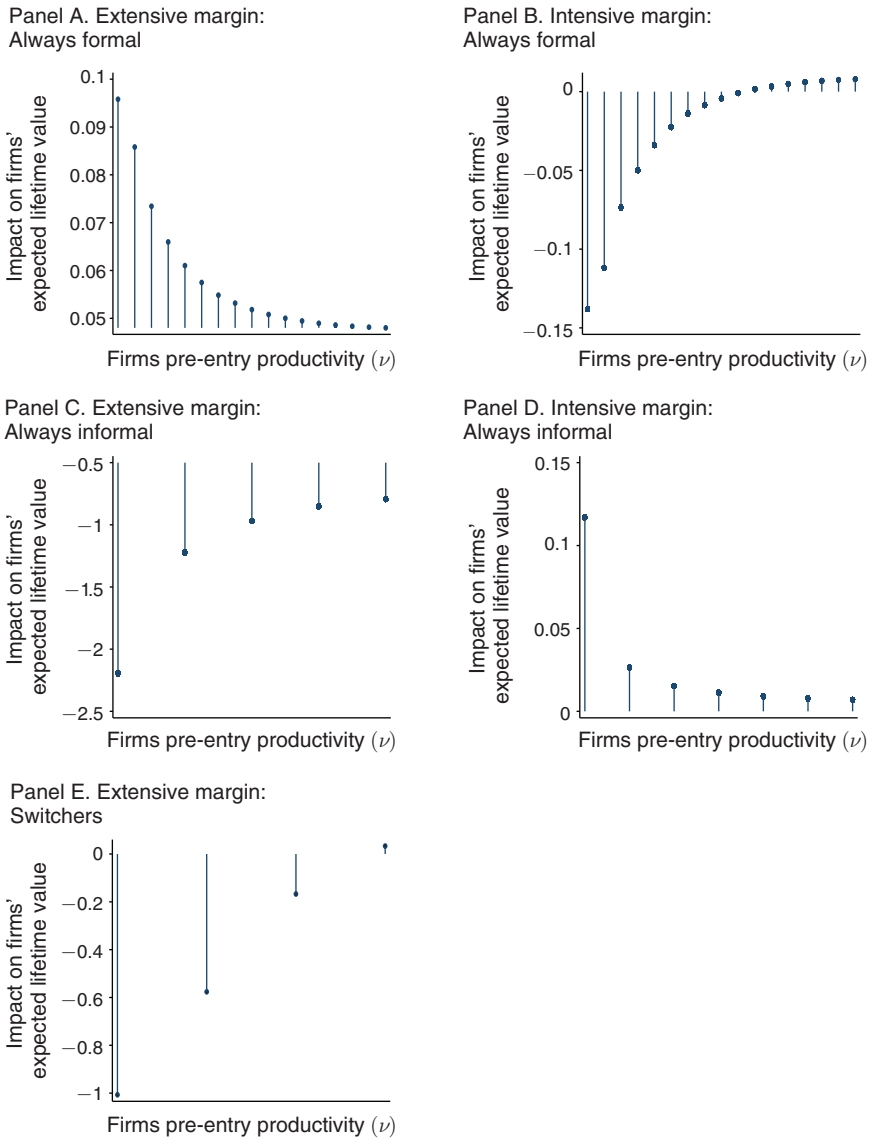

FIGURE 8. PROFILES OF FIRM-LEVEL EFFECTS: INCREASING THE COSTS OF INFORMALITY

for policies (iii) and (iv), which increase the costs of the extensive and intensive margins of informality, respectively.

Panels A and C of Figure 7 show that reducing entry costs hurts high productivity formal firms and all informal firms, except for the switchers (panel E). This negative effect comes from the fact that lowering entry costs induces greater entry into the formal sector, which increases competition and wages for high-skilled workers (Table 6). In contrast, lowering formal sector entry cost greatly benefits low productivity firms in the "always formal" group, as entry costs represent a large

Table 5—The Distribution of Firm-Level Treatment Effects

	All firms	Always formal	Always informal	Switchers
Reducing entry costs				
Mean	−0.103	0.099	−0.172	0.127
Percentile 25	−0.113	0.030	−0.149	0.065
Percentile 50	−0.052	0.093	−0.073	0.137
Percentile 75	−0.025	0.165	−0.044	0.195
Percentile 95	0.192	0.232	−0.028	0.237
Reducing payroll tax				
Mean	−0.525	−0.010	−0.637	−0.124
Percentile 25	−0.629	−0.043	−0.734	−0.140
Percentile 50	−0.333	−0.006	−0.406	−0.124
Percentile 75	−0.192	0.025	−0.260	−0.107
Percentile 95	0.021	0.047	−0.178	−0.091
Higher enforcement: extensive margin				
Mean	−0.934	0.071	−1.638	−0.299
Percentile 25	−1.419	0.058	−1.920	−0.497
Percentile 50	−0.883	0.069	−1.347	−0.224
Percentile 75	0.052	0.083	−1.032	−0.045
Percentile 95	0.087	0.097	−0.829	0.074
Higher enforcement: intensive margin				
Mean	0.038	−0.064	0.055	—
Percentile 25	0.009	−0.102	0.011	—
Percentile 50	0.015	−0.058	0.018	—
Percentile 75	0.033	−0.024	0.039	—
Percentile 95	0.166	0.001	0.189	—

Notes: The firm-level treatment effects are computed using the expression (10) defined in the text. The simulations are the following: in panel A, a reduction in formal sector's entry cost so that $E_f = E_i$; in panel B, a 20 p.p. cut in the payroll tax; in panel C, a reduction in b_i to near zero ($b_i = 0.75$); and panel C, a reduction in b_{f1} and b_{f2} to near zero (0.65 and 0.75, respectively).

fraction of their value relatively to higher productivity firms. This positive effect more than compensates the negative general equilibrium effect (i.e., higher wages). Since the productivity distribution is very skewed to the right, most firms in the "always formal" group perceive a net gain, with an average increase of around 10 percent in their expected net value (Table 5). The switchers also greatly benefit from the reduction in entry costs (panel E): the average increase in firm's expected lifetime value is 13.3 percent, while firms in the ninety-fifth percentile observe an increase as high as 26.7 percent (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 effect 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) comes from the general equilibrium effect on wages: as the payroll tax decreases, the demand for labor increases, leading to a positive effect on equilibrium wages for both skill levels. This wage increase hurts more low productivity firms and resources are shifted away from these firms to more productive ones, which explains the increasing profile observed in panel B. Interestingly, this adverse general equilibrium effect implies that even the switchers are worse off relatively

Table 6—Aggregate Effects

	Baseline	Entry costs	Payroll tax	Extensive mg.	Intensive mg.
Informal labor (*share*)					
All workers	0.351	0.353	0.227	0.224	0.298
Unskilled	0.425	0.425	0.280	0.268	0.357
Skilled	0.266	0.271	0.167	0.173	0.230
Informal firms (share)	0.688	0.435	0.608	0.211	0.741
Informal output (share)	0.199	0.129	0.147	0.034	0.219
Wages					
Skilled	1.000	1.038	1.147	1.009	0.991
Unskilled	1.000	1.004	1.104	0.915	0.998
Skill premium	1.427	1.476	1.482	1.575	1.417
Mass of firms	1.000	1.227	0.866	0.945	0.966
TFP	1.000	0.939	1.087	1.083	1.017
Output	1.000	1.042	1.009	1.032	0.984
Tax revenues	1.000	1.106	0.955	1.222	0.992
Welfare	1.000	1.055	1.044	0.933	1.002

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 welfare measure is given by $U = w\bar{L} + \Pi + T$, where Π denotes total profits in the economy net of total entry costs, $M_f E_f + M_i E_i - M_i = [G(\overline{\nu}_f) - G(\overline{\nu}_i)]M$ and $M_f = [1 - G(\overline{\nu}_f)]M$ denote the measures of entrants into the informal and formal sectors, respectively; and T denotes tax revenues. The simulations are the following: entry costs refers to a reduction in formal sector's entry cost so that $E_f = E_i$; payroll tax refers to a 20 percentage point cut in the payroll tax; extensive mg. refers to a reduction in b_i to near zero ($b_i = 0.75$); and intensive mg. refers to a reduction in b_{f1} and b_{f2} to near zero (0.65 and 0.75, respectively).

to the baseline, even though it is best for them to enter the formal sector in the counterfactual equilibrium. For these firms, the new equilibrium implies an average loss of 13.2 percent in their expected lifetime value (Table 5).

Turning to the experiments that increase the costs of informality, panel A of Figure 8 shows that the firms that are always formal benefit from higher enforcement on the extensive margin of informality, with an average increase of 7.4 percent in their expected lifetime value (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 further reduce their scale in order to remain invisible to the government, which causes an extremely large negative impact on their expected net value (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, intensifying enforcement on this margin of informality substantially increases effective labor costs for these firms. The average effect is a loss of 6.6 percent in firm's lifetime value but with a great degree of heterogeneity, as the effect can be as low as a loss of nearly 15 percent (Figure 8). This policy has a positive impact on informal firms, with an average gain of 5.7 percent. This result comes from the small reduction in equilibrium wages due to a lower demand for labor from low productivity formal firms. For the same reason, high productivity formal firms also observe a slight positive effect.

Economy-Wide Effects.—Reducing formal sector's entry cost leads to a substantial reduction in the share of informal firms, of nearly 25 p.p. (Table 6). The effect on the share of informal workers is however null for both skill levels, 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 basically zero.

The opposite is true when the payroll tax is reduced: informal employment decreases in both skill levels but the share of informal firms does not fall nearly as much. This is observed because the labor tax directly affects formal firms' decision to hire informal or formal labor; 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: it generates a small reduction in the share of informal workers, more so among low-skill workers, but it actually increases 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 these 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 toward informality.

Reducing entry costs also eliminates dead weight losses from wasteful barriers to entry, which substantially increases the mass of active firms in the economy (22.7 percent relatively to the baseline). As a result, competition, production in the formal sector, total output, and wages increase. Since formal firms are more intensive in high-skill labor, the positive effects on wages are concentrated on these workers, which leads to an increase in the skill premium of nearly five percentage points. In addition to greater firm entry, a second channel that causes output 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, 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, regardless of the measure used (share of firms, workers, or output). This generates a positive effect on aggregate TFP (8.3 percent increase) due to composition effects, but at the cost of eliminating low-productivity firms. These two effects generate opposing forces on total output: higher TFP goes in the direction of increasing production, while a lower mass of firms goes in the opposite direction. However, the displacement of low productivity firms is not large and total output increases by 3.2 percent. Since informal firms are more intensive in low-skill labor, these workers suffer a wage loss of 8.5 percent, which leads to an increase in the skill wage premium.

The welfare analysis shows that reducing entry costs leads to a welfare gain of nearly 5.5 percent.[38] This result is partially a consequence of the substantial increase in the mass of active firms, higher output, wages, and tax revenues. Importantly, however, this policy mechanically increases aggregate net profits (which enter directly the welfare measure), as formal sector's entry costs are substantially reduced. This mechanical effect is large: in the partial equilibrium exercise where entry costs are equalized but everything else is kept constant at the baseline level, there is a welfare increase of 6.2 percent. Thus, the general equilibrium effects have a net impact of reducing the welfare gain. This comes from the negative effect that higher wages have on firms, both formal and informal.

Reducing payroll taxes also generates positive welfare effects, of 4.4 percent. This positive effect comes mostly from its positive effect on wages and should be seen as an upper bound, as the labor supply is fixed. Tax revenues are reduced because of the substantial tax cut but are partially compensated by the formalization effects. The substantial increase in wages also leads some firms to exit, as the fixed cost of operation is tied to the equilibrium wage of low-skill workers. As the firms that exit are the least productive ones average TFP increases 8.7 percent, which compensates the reduction in the mass of active firms and therefore total output remains roughly constant.

Higher enforcement on the extensive margin leads to a welfare loss of 6.7 percent. This result is a direct consequence of the fact that the government is now enforcing costly and inefficient regulations to all firms. Even though firms can still evade the payroll tax after formalizing, they must incur heavy entry costs. Even the substantial increase in tax revenues, which are rebated to the households, is not able to compensate this negative effect.[39] Moreover, this policy substantially reduces the equilibrium wage of low-skill workers, which also contributes to the overall negative effect on welfare. It is worth highlighting that this should be seen as an underestimation of the welfare losses from greater enforcement, 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 can be very costly.

This result contrasts with the one found in Meghir, Narita, and Robin (2015), who conclude that increasing enforcement on informal firms leads to a slight unemployment reduction, which seems at odds with the available microevidence (Almeida and Carneiro 2012), and improves welfare. Their framework is not directly comparable to the present one, as they focus on labor market outcomes and use worker-level data (as opposed to focusing on firms and using firm-level data). Nevertheless, one important difference is the absence of fixed costs of entry into the formal sector in their framework. These costs play an important role here,

[38] The welfare measure is the one discussed in Section II: $w\overline{L} + \Pi + T$, where Π denotes total profits in the economy net of total entry costs, $M_f E_f + M_i E_i$, where $M_i = \left[G(\overline{\nu}_f) - G(\overline{\nu}_i)\right] M$ and $M_f = \left[1 - G(\overline{\nu}_f)\right] M$ denote the measures of entrants into the informal and formal sectors, respectively; $w\overline{L}$ is the mass of wages and T denotes tax revenues.

[39] It is worth emphasizing that this effect does not consider the additional revenues that could come from increased fines due to greater enforcement. This positive effect coming from higher tax revenues is closer to the mechanisms highlighted by the literature on fiscal capacity (e.g., Besley and Persson 2013).

as they imply that regulations generate wasteful, fixed costs that are associated to formalization, which is in line with the existing literature (e.g., De Soto 1989; Djankov et al. 2002; Bertrand and Kramarz 2002; Kaplan, Piedra, and Seira 2011; De Mel, McKenzie, and Woodruff 2013). These costs enter directly into the welfare measure and thus play an important role in determining the negative effect from increasing enforcement on the extensive margin. Indeed, if one excludes the cost of entry from the welfare measure, the welfare losses from increasing enforcement disappear.[40]

V. 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 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 a 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 views. The results show that firms that are potentially productive and that formalize and succeed when formal sector's entry costs are removed constitute a small fraction of all informal firms (11.5 percent). The view that argues that informal firms choose informality to exploit the cost advantages of noncompliance even though they are productive enough to survive in the formal sector corresponds to a large fraction of all informal firms, 35.9 percent. The remaining firms correspond to those too unproductive to ever become formal.

The counterfactual exercises show that there are winners and losers in all policies and that there is substantial heterogeneity in policy effects between groups (switchers, always formal, and always informal firms) and within groups. At the aggregate level, I find that increasing enforcement is highly effective in reducing informality but it reduces welfare in the economy. Reducing formal sector's entry costs is not as effective in reducing informality but generates 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.

[40] Ulyssea (2010) develops a simpler matching model than the one in Meghir, Narita, and Robin (2015) but that includes a fixed cost of entry into the formal sector that comes from regulatory costs. The author finds that increasing enforcement on informal firms reduces welfare and leads to higher unemployment.

REFERENCES

Adda, Jérôme, and Russell W. Cooper. 2003. *Dynamic Economics: Quantitative Methods and Applications.* Cambridge, MA: MIT Press.

Adda, Jérôme, Christian Dustmann, and Katrien Stevens. 2017. "The Career Costs of Children." *Journal of Political Economy* 125 (2): 293–337.

Allingham, Michael G., and Agnar Sandmo. 1972. "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics* 1 (3–4): 323–38.

Almeida, Rita, and Pedro Carneiro. 2009. "Enforcement of Labor Regulation and Firm Size." *Journal of Comparative Economics* 37 (1): 28–46.

Almeida, Rita, and Pedro Carneiro. 2012. "Enforcement of Labor Regulation and Informality." *American Economic Journal: Applied Economics* 4 (3): 64–89.

Altonji, Joseph G., and Lewis M. Segal. 1996. "Small-Sample Bias in GMM Estimation of Covariance Structures." *Journal of Business & Economic Statistics* 14 (3): 353–66.

Amaral, Pedro S., and Erwan Quintin. 2006. "A Competitive Model of the Informal Sector." *Journal of Monetary Economics* 53 (7): 1541–53.

Bertrand, Marianne, Chang-Tai Hsieh, and Nick Tsivanidis. 2015. "Contract Labor and Firm Growth in India." Unpublished.

Bertrand, Marianne, and Francis Kramarz. 2002. "Does Entry Regulation Hinder Job Creation? Evidence from the French Retail Industry." *Quarterly Journal of Economics* 117 (4): 1369–1413.

Besley, Timothy, and Torsten Persson. 2013. "Taxation and Development." In *Handbook of Public Economics,* Vol. 5, edited by Alan J. Auerbach et al., 51–110. Amsterdam: Elsevier.

Bosch, Mariano, and Julen Esteban-Pretel. 2012. "Job Creation and Job Destruction in the Presence of Informal Markets." *Journal of Development Economics* 98 (2): 270–86.

Bruhn, Miriam. 2011. "License to Sell: The Effect of Business Registration Reform on Entrepreneurial Activity in Mexico." *Review of Economics and Statistics* 93 (1): 382–86.

Busso, Matías, María Victoria Fazio, and Santiago Levy. 2012. "(In)Formal and (Un)Productive: The Productivity Costs of Excessive Informality in Mexico." IDB Working Paper IDB-WP-341.

Charlot, Olivier, Franck Malherbet, and Cristina Terra. 2015. "Informality in Developing Economies: Regulation and Fiscal Policies." *Journal of Economic Dynamics and Control* 51: 1–27.

Colombi, Roberto. 1990. "A New Model of Income Distribution: The Pareto-Lognormal Distribution." In *Income and Wealth Distribution, Inequality and Poverty*, edited by Camilo Dagum and Michele Zenga, 18–32. Berlin: Springer-Verlag.

De Andrade, Gustavo Henrique, Miriam Bruhn, and David J. McKenzie. 2014. "A Helping Hand or the Long Arm of the Law? Experimental Evidence on What Governments Can Do to Formalize Firms." *World Bank Economic Review* 30 (1): 24–54.

de la Parra, Brenda Samaniego. 2016. "Formal Firms, Informal Workers and Household Labor Supply in Mexico." Unpublished.

de Mel, Suresh, David McKenzie, and Christopher Woodruff. 2013. "The Demand for, and Consequences of, Formalization among Informal Firms in Sri Lanka." *American Economic Journal: Applied Economics* 5 (2): 122–50.

de Paula, Áureo, and José A. Scheinkman. 2010. "Value-Added Taxes, Chain Effects, and Informality." *American Economic Journal: Macroeconomics* 2 (4): 195–221.

de Paula, Áureo, and José A. Scheinkman. 2011. "The Informal Sector: An Equilibrium Model and Some Empirical Evidence from Brazil." *Review of Income and Wealth* 57 (S1): S8–S26.

D'Erasmo, Pablo N., and Herman J. Moscoso Boedo. 2012. "Financial Structure, Informality and Development." *Journal of Monetary Economics* 59 (3): 286–302.

De Soto, Hernando. 1989. *The Other Path: The Economic Answer to Terrorism.* New York: Harper & Row.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-De-Silanes, and Andrei Shleifer. 2002. "The Regulation of Entry." *Quarterly Journal of Economics* 117 (1): 1–37.

Fajnzylber, Pablo, William F. Maloney, and Gabriel V. Montes-Rojas. 2011. "Does Formality Improve Micro-Firm Performance? Evidence from the Brazilian SIMPLES Program." *Journal of Development Economics* 94 (2): 262–76.

Farrell, Diana. 2004. "The Hidden Dangers of the Informal Economy." *McKinsey Quarterly* 26–37.

Fortin, Bernard, Nicolas Marceau, and Luc Savard. 1997. "Taxation, Wage Controls and the Informal Sector." *Journal of Public Economics* 66 (2): 293–312.

Garicano, Luis, Claire Lelarge, and John van Reenen. 2016. "Firm Size Distortions and the Productivity Distribution: Evidence from France." *American Economic Review* 106 (11): 3439–79.

Gasparini, Leonardo, and Leopoldo Tornarolli. 2009. "Labor Informality in Latin America and the Caribbean: Patterns and Trends from Household Survey Microdata." *Desarrollo y Sociedad* (63): 13–80.

Gouriéroux, Christian, and Alain Monfort. 1996. *Simulation-Based Econometric Methods*. Oxford, UK: Oxford University Press.

Guner, Nezih, Gustavo Ventura, and Yi Xu. 2008. "Macroeconomic Implications of Size-Dependent Policies." *Review of Economic Dynamics* 11 (4): 721–44.

Heckman, James J. 2001. "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture." *Journal of Political Economy* 109 (4): 673–748.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124 (4): 1403–48.

Hsieh, Chang-Tai, and Benjamin A. Olken. 2014. "The Missing 'Missing Middle.'" Unpublished.

Ichimura, Hidehiko, and Christopher Taber. 2002. "Semiparametric Reduced-Form Estimation of Tuition Subsidies." *American Economic Review* 92 (2): 286–92.

Kaplan, David S., Eduardo Piedra, and Enrique Seira. 2011. "Entry Regulation and Business Start-Ups: Evidence from Mexico." *Journal of Public Economics* 95 (11–12): 1501–15.

Keane, Michael P., Petra E. Todd, and Kenneth Wolpin. 2011. "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications." In *Handbook of Labor Economics*, Vol. 4A, edited by Orley Ashenfelter and David Card, 331–461. New York: Elsevier.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. 2011. "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark." *Econometrica* 79 (3): 651–92.

Kuehn, Zoë. 2014. "Tax Rates, Governance, and the Informal Economy in High Income Countries." *Economic Inquiry* 52 (1): 405–30.

La Porta, Rafael, and Andrei Shleifer. 2008. "The Unofficial Economy and Economic Development." *Brooking Papers on Economic Activity* (Fall): 275–352.

La Porta, Rafael, and Andrei Shleifer. 2014. "Informality and Development." *Journal of Economic Perspectives* 28 (3): 109–26.

Leal Ordóñez, Julio César. 2014. "Tax Collection, the Informal Sector, and Productivity." *Review of Economic Dynamics* 17 (2): 262–86.

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

Luttmer, Erzo G. J. 2007. "Selection, Growth and the Size Distribution of Firms." *Quarterly Journal of Economics* 122 (3): 1103–44.

Maloney, William F. 2004. "Informality Revisited." *World Development* 32 (7): 1159–78.

Meghir, Costas, Renata Narita, and Jean-Marc Robin. 2015. "Wages and Informality in Developing Countries." *American Economic Review* 105 (4): 1509–46.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.

Monteiro, Joana C. M., and Juliano 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): 105–15.

Perry, Guillermo E., William F. Maloney, Omar S. Arias, Pablo Fajnzylber, Andrew D. Mason, and Jaime Saavedra-Chanduvi. 2007. *Informality: Exit and Exclusion*. Washington, DC: The World Bank.

Prado, Mauricio. 2011. "Government Policy in the Formal and Informal Sectors." *European Economic Review* 55 (8): 1120–36.

Rauch, James E. 1991. "Modeling the Informal Sector Formally." *Journal of Development Economics* 35 (1): 33–47.

Rocha, Rudi, Gabriel Ulyssea, and Laísa Rachter. 2018. "Do Lower Taxes Reduce Informality? Evidence from Brazil." *Journal of Development Economics* 134: 28–49.

Rothschild, Casey, and Florian Scheuer. 2016. "Optimal Taxation with Rent-Seeking." *Review of Economic Studies* 83 (3): 1225–62.

Slemrod, Joel, and Shlomo Yitzhaki. 2002. "Tax Avoidance, Evasion, and Administration." In *Handbook of Public Economics,* Vol. 3, edited by Martin Feldstein and Alan J. Auerbach, 1423–70. Amsterdam: North-Holland.

Ulyssea, Gabriel. 2010. "Regulation of Entry, Labor Market Institutions and the Informal Sector." *Journal of Development Economics* 91 (1): 87–99.

Ulyssea, Gabriel. 2018. "Firms, Informality, and Development: Theory and Evidence from Brazil: Dataset." *American Economic Review.* https://doi.org/10.1257/aer.20141745.