The Effects of Tax Enforcement on the Firm Size Distribution and Aggregate Productivity*

Nicola Corbellini

University of Minnesota

July 20, 2024

Click here for the most recent version

Abstract

How does tax enforcement affect the firm size distribution and total factor productivity (TFP)? To answer this question, I develop a quantitative model characterized by heterogeneous agents who choose whether to be workers, entrepreneurs in the formal sector, or entrepreneurs in the informal sector. Informal entrepreneurs do not pay taxes but face a probability of detection that is increasing in firm size. In the model, stricter tax enforcement results in lower informality and affects the firm size distribution and TFP through two mechanisms: as tax enforcement becomes stricter, fewer relatively unproductive agents choose to be entrepreneurs, and fewer entrepreneurs choose to operate in the informal sector. Using data from Brazil, I calibrate the model and estimate that a counterfactual tax enforcement that reduces the informality rate from 36% to 30% of total output—the value measured in the weighted average of the six largest Latin American economies—would account for about 9% and 28% of the observed differences in TFP and average firm size.

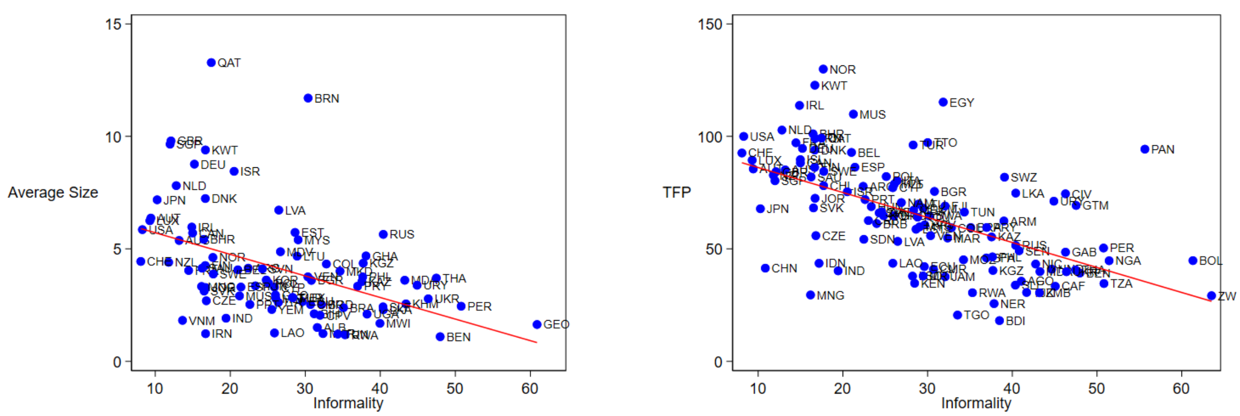
*I am grateful to my advisor V. V. Chari for his encouragement and insightful discussions throughout the development of this paper. I also thank Larry Jones and Christopher Phelan for their excellent comments and suggestions. I received useful feedback from Ricardo Alves Monteiro, Diego Ascarza-Mendoza, Mauricio Barbosa Alves, Braulio Britos, Loukas Karabarbounis, and participants in the Minnesota Workshop in Growth and Development. I benefited from the publicly available Gabriel Ulyssea's replication files for the computation of the moments used in the calibration. I also benefitted from the Data Zoom project for accessing Brazilian microdata. All remaining errors are mine.

1 Introduction

Developing countries are characterized by the presence of a firm size distribution that is highly concentrated on businesses with a low number of employees. In fact, an overwhelming majority of firms and a relatively large fraction of value added are accounted for by micro firms with at most 5 employees.[1] Informality rates among small firms are substantial, reflecting the government's inability to detect and tax these entities. For instance, the informality rate among firms smaller than 5 employees is 86% in Brazil and 94% in Mexico[2]. In addition, compared to advanced economies, developing countries display lower measures of total factor productivity (TFP), which is generally deemed as the main source of cross-country differences in per-capita income. Figure 1 shows cross-country correlations between informality rates, average firm size (panel a), and TFP (panel b). To sum up, developing countries display coexisting high informality and tax evasion rates, low average firm size, and low aggregate productivity.[3]

Figure 1: Correlation between Informality, Average Firm Size, and TFP across Countries.

(a) Informality and Average Firm Size. (b) Informality and TFP.



Source: Elgin et al. (2021) for informality, Bento and Restuccia (2021) for average size, Feenstra et al. (2015) for TFP. Informality and TFP data are country averages over the period 2005-2015.

How much does tax enforcement account for the observed differences in size and productivity across countries? This paper investigates the impact of tax enforcement on the firm size

[1]Whenever the number of employees is mentioned in this paper, it includes the owner.
[2]Data for Brazil are presented in Section 3.1. See Alvarez and Ruane (2023) for Mexican data.
[3]Appendix A.1 provides a cross-country analysis of the relationship between informality and TFP.

distribution and TFP. To assess the magnitude of the impact, I build a dynamic quantitative model characterized by heterogeneous agents, occupational choice, and the presence of an informal sector. Agents can choose whether to be workers, entrepreneurs in the formal sector, or entrepreneurs in the informal sector. Formal entrepreneurs pay sales and payroll taxes, while informal entrepreneurs do not. However, the latter are detected with a probability that increases with the size of their business. In case of detection, informal entrepreneurs pay the due amount of taxes plus a proportional fine. Tax enforcement is then summarized by the combination of tax rates, penalty rates, and a probability of detection that is an increasing function of the number of employees in the business.

In the model, stricter tax enforcement affects the firm size distribution and TFP through two mechanisms. First, the profitability of operating in the informal sector declines, inducing some relatively unproductive agents to be workers. These agents would operate as informal entrepreneurs with looser tax enforcement because they would be able to circumvent tax payments. Moreover, these entrepreneurs would operate at a low scale, given the positive relationship between productivity and size. Consequently, average size and productivity increase thanks to a composition effect, referred to as the *selection channel*. Second, to avoid detection by the tax authority, informal entrepreneurs operate at a lower scale compared to formal entrepreneurs of the same productivity. Therefore, as the rate of informality decreases due to stricter tax enforcement, fewer entrepreneurs decide to keep their scale at suboptimally low levels. This *detection channel* has a positive effect on average firm size and allocative efficiency.

To calibrate the model, I use data from Brazil. More specifically, I substantially rely on the *Economia Informal Urbana* (ECINF), a detailed firm-level survey lastly conducted in 2003, which contains information on informality status and number of employees. The calibrated model successfully replicates several features of the Brazilian economy, such as the high share of micro and small firms, high informality (36% of GDP) and tax evasion rates, high entrepreneurship and self-employment rates, and the observation that formal entrepreneurs are, on average, larger in size and more productive than informal entrepreneurs.

I then perform a counterfactual analysis to measure the impact of tax enforcement. To simulate stricter tax enforcement in Brazil, I change the parametrization of the detection probability function to match an informality rate of 30%, which is the GDP-weighted average of the six largest Latin American economies. The counterfactual calibrated economy generates an increase in output and TFP by approximately 4.3% and 0.5%, respectively, and a rise in average firm size from 2.08 in the baseline to 2.37. These changes account for about 13%, 9%, and 28% of the observed differences in GDP per capita, TFP, and average firm size between Brazil and the weighted averages of the countries in the group. These results are consistent across several exercises in which different countries or groups of countries are taken as counterfactuals. Overall, tax enforcement can account for about 25-30% of the estimated differences in average firm size and for about 8-15% of the estimated differences in GDP per capita and TFP.

What are the mechanisms through which stricter tax enforcement results in higher TFP and average firm size? The counterfactual economy displays a lower rate of entrepreneurship (3 percentage points lower than in the baseline) and lower rates of informality for given values of firm-level productivity. Overall, there is a reallocation of resources from low-productivity informal entrepreneurs to high-productivity formal entrepreneurs. From a quantitative point of view, decomposing the impact between the selection and detection channel highlights that both channels have substantial positive impacts on average firm size (they both account for about 50% of the overall impact), while only the detection channel has a significant positive effect on TFP.

This paper proceeds as follows. Section 2 introduces the quantitative model. Section 3 presents the data and the baseline calibration of the model. Section 4 describes the counterfactual experiments and presents the quantitative results. Section 5 concludes.

Related literature. This paper relates to three main strands of literature. First, it relates to the literature studying the aggregate effects of tax enforcement, either in developing countries or in advanced economies. My framework shows some similarities with the work of Leal Ordóñez (2014), who focuses on Mexico, and Di Nola et al. (2021), who focus on the US, for the presence of selection and detection channels as the main mechanisms affecting aggregate outcomes. With regard to cross-country studies, Bachas et al. (2019) analyze how size-dependent taxation affects TFP across several countries and sectors. Differently from this paper, which relies on informal businesses' statistics for the quantification, Bachas et al. (2019) use World Bank data from registered formal firms to estimate the model and perform counterfactual analysis.

The paper is also related to the literature that analyzes the relationship between informality and aggregate outcomes. These studies often focus on a single country such as Brazil (Ulyssea (2018), Erosa et al. (2023)) or Mexico (Leal Ordóñez (2014)). Ulyssea (2018) studies the effects of policy reforms reducing informality on output, TFP, and welfare. The author develops a quantitative model that features two margins of informality: an *extensive margin*, associated with the business side, and an *intensive margin*, which reflects the hiring of "off-the-books" workers by formal firms. He then shows that policies affecting the intensive margin do not necessarily increase output and welfare. Erosa et al. (2023) extend Ulyssea's framework to assess the interaction between informality and financial frictions. While getting similar qualitative results as in Ulyssea (2018), they show that tight borrowing constraints amplify the negative effects of informality on the allocation of resources. My work distinguishes from the above papers in three dimensions. First, my model introduces a microfoundation of the relationship between size and the cost of informality that depends on tax enforcement and is disciplined by statutory tax rates and moments from the informal firm size distribution. Second, I focus my analysis on the effects on average firm size and TFP. Third, I estimate how tax enforcement can account for the differences in average firm size and TFP between Brazil and other countries displaying lower informality rates.

Finally, the paper is connected with the literature studying the effects of correlated distortions and size-dependent policies on TFP (Restuccia and Rogerson (2008); Guner et al. (2008); Hsieh and Klenow (2009); Buera et al. (2013)) and the firm size distribution (Garicano et al. (2016); Bento and Restuccia (2017); Bento and Restuccia (2021)). For example, Bento and Restuccia (2021) show how different levels of elasticity between distortions and size can account for cross-country differences in average establishment size. My contribution to this literature is to highlight how tax enforcement could represent one of the underlying sources of the measured differences in correlated distortions across countries and a relevant factor in accounting for differences in aggregate outcomes.

2 Model

Time is discrete. I consider a small open economy with fixed interest rate r. The economy is populated by a continuum of infinitely-lived individuals who are heterogeneous in their ability to run a business. Individuals choose whether to be entrepreneurs or workers. Each worker inelastically supplies one unit of labor. Entrepreneurs additionally choose whether to operate the firm in the formal or informal sector.

Entrepreneurs. Both formal and informal entrepreneurs adopt the same technology in producing a homogeneous final good. Entrepreneur i's production function is

$$y_i = z_i \times \underbrace{(k_i^{\alpha}(\bar{\ell} + n_i)^{1-\alpha})^{\gamma}}_{f(k_i,n_i)} - \theta n_i,$$

where y denotes output, z is managerial ability (i.e., firm-level productivity), k is capital, $\bar{\ell}$ is the owner's labor supply entering the production function, and n is hired labor. $\alpha \in (0,1)$ denotes the capital share, while $\gamma \in (0,1)$ denotes the span-of-control parameter. The value of $\bar{\ell}$ is common across entrepreneurs. Entrepreneurial ability z_i evolves over time according to an $AR(1)$ process. Capital is borrowed at a rate $(r + \delta)$, where δ denotes depreciation, while labor is provided by workers at an hourly wage rate w. Entrepreneurs also pay a variable cost proportional to the number of employees $\theta \times n_i$, which reflects the cost of managing employees.[4] The baseline economy is characterized by perfect financial markets, so entrepreneurs can rent any desired amount of capital for production.

Each entrepreneur decides whether to keep the firm formal or informal. Formal firms pay an output (sales) tax $\tau_y z_i f(k_i, n_i)$ and a payroll tax $\tau_n wn$, while informal firms do not. However, informal firms are detected with probability $p(n)$, with $p'(n) > 0$, that is, the detection probability is increasing in the amount of hired labor used for production. A detected firm has to pay a surcharge $(1 + \kappa)$ of its due taxes to the government, with $\kappa > 0$. Therefore, tax enforcement is defined by the combination of tax rates τ_y and τ_n, a surcharge rate κ, and a probability of detection function $p(n)$.

[4] Adding the variable cost $\theta \times n$ helps capture the very high level of self-employment among Brazilian informal firms.

Preferences. Agents maximize the stream of per-period utilities and discount the future at rate β:[5]

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t),$$

where c_t denotes consumption and the utility function $u(\cdot)$ satisfies the usual properties.

Government. The government balances the budget in every period. It collects taxes from entrepreneurs in the formal sector and ex-post punishment amounts from entrepreneurs in the informal sector. The collected revenues are used to finance an exogenous amount of public expenditures G, which does not enter the agents' utility function.

Timing. Within each period, timing is the following:

(i) Productivity shocks realize.

(ii) Agents decide whether to be workers, formal, or informal entrepreneurs.

(iii) Capital and labor decisions are made. Production occurs.

(iv) Formal firms pay taxes to the government. Detected informal firms pay the ex-post punishment amount.

(v) Individuals make consumption and saving decisions given after-tax (or punishment) earnings.

Recursive Problems. The worker's problem can be written in the following recursive form:

$$\begin{aligned} V_W(z,a) = &\max_{c,a'} \{u(c) + \beta \mathbb{E}_{z'} V(z',a')\} \\ s.t.\ \ & c + a' \leq w + (1+r)a, \\ & a' \geq 0. \end{aligned} \tag{1}$$

[5] I remove the i subscripts hereafter for ease of notation.

The formal entrepreneur's problem can be written in the following recursive form:

$$\begin{aligned} V_F(z,a) = &\max_{c,a',k,n} \{u(c) + \beta \mathbb{E}_{z'} V(z',a')\} \\ s.t.\ \ & c + a' \leq \pi_F(z) + (1+r)a, \\ & \pi_F(z) = (1-\tau_y) z (k^{\alpha} (\bar{\ell} + n)^{1-\alpha})^{\gamma} - (r+\delta)k - [w(1+\tau_n) + \theta] n \\ & a' \geq 0. \end{aligned} \tag{2}$$

The informal entrepreneur's value function is

$$V_I(z,a) = \max_{k,n} \{(1-p(n)) V_I^{nd}(z,a,k,n) + p(n) V_I^{d}(z,a,k,n)\}, \tag{3}$$

where $V_I^d(z,a,k,n)$ and $V_I^{nd}(z,a,k,n)$ denote the value functions of detected and not detected entrepreneurs, respectively.
For $j = \{d, nd\}$, the recursive problem can be written as

$$\begin{aligned} V_I^j(z,a,k,n) = &\max_{c,a'} \{u(c) + \beta \mathbb{E}_{z'} V(z',a'))\} \\ s.t.\ \ c + a' \leq &\pi_I^j(z) + (1+r)a, \\ \pi_I^j(z) = &z (k^{\alpha} (\bar{\ell} + n)^{1-\alpha})^{\gamma} - (r+\delta)k - (w+\theta)n - \\ & \mathbb{1}_{j=d} \times [(1+\kappa)(\tau_y z (k^{\alpha} (\bar{\ell}+n)^{1-\alpha})^{\gamma} + \tau_n w n)], \\ a' \geq &0. \end{aligned}$$

Finally, an agent with productivity z and assets a chooses an occupation according to

$$V(z,a) = \max\{V_W(z,a), V_I(z,a), V_F(z,a)\}. \tag{4}$$

Equilibrium. Given an interest rate r, government expenditures G, a government policy

$\{\tau_y, \tau_n, \kappa\}$, and a detection probability function $p(\cdot)$, a *stationary competitive equilibrium* is wage w, occupational choices $o(z,a)$, entrepreneurs' capital and labor policy functions $k(z,a)$, $n(z,a)$, consumption and asset allocations for workers, formal entrepreneurs, detected informal entrepreneurs, and undetected informal entrepreneurs $c_W(z,a)$, $c_F(z,a)$, $c_{I,d}(z,a)$, $c_{I,nd}(z,a)$, $a_W(z,a)$, $a_F(z,a)$, $a_{I,d}(z,a)$, $a_{I,nd}(z,a)$, value functions $V(z,a)$, $V_W(z,a)$, $V_F(z,a)$, $V_I(z,a)$, and a distribution $\mu(z,a)$, such that:

- Worker's value function $V_W(z,a)$ and allocations $c_W(z,a)$, $a_W(z,a)$ solve problem (1).
- Formal entrepreneur's value function $V_F(z,a)$ and allocations $c_F(z,a)$, $a_F(z,a)$, $k(z,a)$, $n(z,a)$ solve problem (2).
- Informal entrepreneur's value function $V_I(z,a)$ and allocations $c_{I,d}(z,a)$, $c_{I,nd}(z,a)$, $a_{I,d}(z,a)$, $a_{I,nd}(z,a)$, $k(z,a)$, $n(z,a)$ solve problem (3).
- Occupational choices $o(z,a)$ are consistent with (4).
- The labor market clears:

$$L = \int_{o(z,a)=W} d\mu(z,a) = \int n(z,a) d\mu(z,a).$$

- Government balances its budget:

$$\begin{aligned} G = &\int_{o(z,a)=F} \big(\tau_y z f(k(z,a), n(z,a)) + \tau_n w n(z,a)\big) \times d\mu(z,a) \\ &+ \int_{o(z,a)=I} p(n(z,a)) \times (1+\kappa)\big(\tau_y z f(k(z,a), n(z,a)) + \tau_n w n(z,a)\big) \times d\mu(z,a). \end{aligned}$$

- The current account surplus/deficit is stationary:

$$\begin{aligned} CA = &\int z f(k(z,a), n(z,a)) d\mu(z,a) - \delta K - G - \theta L \\ &- \int_{o(z,a)=W} c_W(z,a) d\mu(z,a) - \int_{o(z,a)=F} c_F(z,a) d\mu(z,a) \\ &- \int_{o(z,a)=I} \Big[c_{I,d}(z,a) p(n(z,a)) + c_{I,nd}(z,a)(1 - p(n(z,a))) \Big] d\mu(z,a). \end{aligned}$$

- The distribution $\mu(z, a)$ is stationary. Given a one-period ahead transition operator $\mathcal{M}$,

$$\mu = \mathcal{M}(\mu).$$

3 Calibration

In this section, I first describe the Brazilian data used for the calibration (Section 3.1). Then, I introduce the functional forms (Section 3.2) and I separate model parameters in two groups: those who are externally set based on conventional or statutory values, and those who are internally calibrated. For the latter, I choose the parametrization that minimizes the distance between model and data moments (Section 3.3). Finally, I show how the calibrated model performs along targeted and non-targeted dimensions (Section 3.4).

3.1 Data

Data for Brazil are collected from different sources. Moments of informal firms are collected from the *Economia Informal Urbana* (ECINF), a representative survey that was lastly conducted in 2003 to collect information on the informal sector in Brazil. The survey contains characteristics of self-employed workers and employers running a non-agricultural business with up to five employees. Owners are classified as informal if they do not possess a tax identification number (*Cadastro Nacional de Pessoa Juridica* (CNPJ)). By matching owners with employees, it is possible to obtain the number of employees for each business. Additional information such as sector and financial variables is also available. The final sample from which I compute informal firm size distribution moments used in the calibration contains about 30,000 firms.[6] Table 1 summarizes sector composition and size distribution by formality status.

While ECINF displays a comprehensive picture of small formal and informal firms, the size threshold of 5 employees makes the sample not representative of the universe of formal firms.

[6]Appendix A.5 provides more details on how I obtain the samples from which I compute the moments used in the calibration.

Table 1: ECINF. Sector and Size Statistics.

	Formal Shares	Informal Shares	Total Shares
Industry			
Manufacturing	8.69	11.27	10.82
Retail	50.50	37.70	39.93
Services	40.81	51.03	49.25
Number of employees			
1	28.95	79.30	70.53
2	27.39	14.14	16.45
3	16.35	3.68	5.89
4	11.71	1.72	3.46
5	8.65	0.79	2.16
6	4.49	0.30	1.03
7	2.47	0.07	0.49
Observations (#)	5,261	24,924	30,185

Source: Own calculations based on ECINF (2003). Note that shares reflect the raw unweighted sample.

Therefore, I integrate ECINF with statistics on formal firms taken from the *Relação Anual de Informações Sociais* (RAIS), an administrative dataset containing number and characteristics of workers as reported by formal employers.[7]

I use data from the 2003 *Pesquisa Nacional por Amostra de Domicílios* (PNAD), a repeated cross-section representative at the national level, to compute the rate of entrepreneurship. Individuals who are either self-employed or employers are classified as entrepreneurs.

Finally, the share of informal output in the Brazilian economy is taken from Elgin et al. (2021), which provides estimates of informality rates for more than 160 countries over the period 1990-2019.[8] The value for Brazil in 2003 is 36%.

Table 2 summarizes the main data. Of particular relevance are the high rate of entrepreneurship (32%), the high share of informal firms (76%), and the fact that 95% of informal firms have at most two employees.

[7] I do not have direct access to the RAIS microdata since it is a restricted administrative dataset. However, I can obtain firm size distribution moments from tabular data and from other papers' statistics (e.g., Ulyssea (2018), Erosa et al. (2023)).

[8] These estimates are obtained through national account statistics and the use of a dynamic general equilibrium model, which determines how optimizing households will allocate labor between formal and informal sectors. See Appendix A.1 for more details.

Table 2: Data for Brazil in 2003.

Moments	Source	Data
Informality		
Informality/GDP	Elgin et al. (2021)	0.36
Share of informal firms[9]	Erosa et al. (2023)	0.76
Entrepreneurship rate	PNAD	0.32
Size distribution (formal) : firms shares		
1-5	RAIS	0.70
6-10	RAIS	0.14
11-20	RAIS	0.08
21-50	RAIS	0.05
Size distribution (informal) : firms shares		
≤ 1	ECINF	0.82
≤ 2	ECINF	0.95
≤ 5	ECINF	0.998

Source: Own calculations based on ECINF and PNAD (2003), tabular data from RAIS.

3.2 Functional Forms

I set the utility function to be $u(c_t) = log(c_t)$.

Managerial productivity evolves according to the following AR(1) process:

$$log(z_{t+1}) = \rho \, log(z_t) + \nu_{t+1}, \qquad \nu \sim N(0, \sigma_\nu^2).$$

Where ρ and σ_ν denote the persistence and the standard deviation of the productivity process.

Regarding the detection probability function, I follow Di Nola et al. (2021) in adopting a logistic function:

$$p(n) = \frac{1}{1 + p_1 exp(-p_2 n)}.$$

The two parameters to be estimated are p_1 and p_2, which control the inflection point and the slope of the detection probability function, respectively.

[9]This number is obtained by merging sample data from ECINF with administrative data from RAIS.

Occupational taste shocks. To smooth aggregate labor demand and supply functions and ease the convergence to an equilibrium, I introduce individual-level occupational taste shocks which are realized at the same time as productivity shocks.[10] These shocks are (i) distributed according to a Gumbel distribution with scale parameter σ_ϵ, (ii) iid across occupational choices within individuals, and (iii) iid across individuals and across time. I set the parameter σ_ϵ to 0.1.

The occupational choice for an agent with productivity z, assets a, and taste shocks $\epsilon_W, \epsilon_I, \epsilon_F$ is consistent with

$$V(z,a) = \max\{V_W(z,a) + \epsilon_W, V_I(z,a) + \epsilon_I, V_F(z,a) + \epsilon_F\}.$$

Since the shocks are iid and follow a Gumbel distribution with scale parameter σ_ϵ, the (ex-ante) probability of choosing occupation $o \in \{W, I, F\}$ given (z,a) is

$$q(o|z,a) = exp\left(\frac{V_o(z,a)}{\sigma_\epsilon}\right) \Bigg/ \sum_{o' \in \{W,I,F\}} exp\left(\frac{V_{o'}(z,a)}{\sigma_\epsilon}\right).$$

Moreover, the value function $V(z,a)$ is given by the log-sum formula

$$V(z,a) = \sigma_\epsilon \; log\left(\sum_{o' \in \{W,I,F\}} exp\left(\frac{V_{o'}(z,a)}{\sigma_\epsilon}\right)\right).$$

3.3 Choice of Parameters

Externally set parameters. Table 3 shows the parameters that are calibrated outside the model. I set the discount factor β to .96, the capital depreciation rate δ to .05, the capital share α to .33, and the span-of-control parameter ν to .9. Tax and surcharge rates are set according to their statutory values. The sales tax τ_y is set to .293, while the payroll tax is set to .375. The output tax includes two VAT federal taxes of 20 and 9 percent, respectively, while the payroll tax includes social security contributions (20 percent), direct payroll tax (9 percent), and severance contributions (8.5 percent).[11] Finally, the surcharge rate κ is set

[10] The sole purpose of introducing these shocks is computational ease, and the main results do not depend on their inclusion.

[11] See Ulyssea (2018) for more details.

to the regular penalty, which is 75% of the due taxes.[12]

Table 3: Externally Set Parameters.

Parameter	**Description**	**Value**	**Source**
β	Discount factor	0.96	Standard value
δ	Capital depreciation rate	0.05	Standard value
α	Capital share	0.33	Standard value
γ	Span-of-control	0.9	Standard value
τ_y	Sales tax	29.3%	Statutory rate
τ_n	Payroll tax	37.5%	Statutory rate
κ	Surcharge rate	75%	Statutory rate

Internally calibrated parameters. There are 6 parameters left to be estimated: the persistence and the standard deviation of the productivity shock ρ and σ_ν, the detection probability parameters p_1 and p_2, the entrepreneurial amount of labor entering the production function $\bar{\ell}$, and the variable cost parameter θ. Model parameters are jointly calibrated and chosen to minimize the distance between the following model statistics and their corresponding moments in the data:

1. Informality rate (output): share of output produced by informal entrepreneurs.
2. Informality rate (firms): share of informal firms (out of all firms).
3. Share of informal self-employed (out of informal firms).
4. Share of informal firms with less than 2 employees (out of informal firms).
5. Share of formal firms with less than 5 employees (out of formal firms).
6. Entrepreneurship rate: sum of self-employed and employers (out of the labor force).

It is important to highlight that parameters p_1 and p_2 are mostly informative of the share of informal output and informal firms in the economy, while ρ and σ_ν mainly affect the firm size distribution moments. On the other hand, $\bar{\ell}$ and θ mainly determine the entrepreneurship and self-employment rates. Table 4 shows the most informative moment for each internally calibrated parameter.

[12] See Franjo et al. (2022).

Table 4: Internally Calibrated Parameters.

Parameter	**Description**	**Moment**
p_1	Detection probability (intercept)	Informality rate (output)
p_2	Detection probability (slope)	Informality rate (firms)
$\bar{\ell}$	Owner labor	Entrepreneurship rate
θ	Variable cost of employees	Share informal firms ≤ 1
ρ	Persistence shock	Share informal firms ≤ 2
σ_ν	Standard deviation shock	Share formal firms ≤ 5

3.4 Calibration Results

Model fit. Table 5 shows the model fit over targeted dimensions and the associated parameter values. The model replicates well the share of informal output, the entrepreneurship rate and the informal firm size distribution moments, while it slightly underestimates the share of small (at most 5 employees) formal firms (62% in the model compared to 70% in the data) and it overestimates the share of informal entrepreneurs (84% in the model compared to 76% in the data).

Table 5: Data and Model Statistics over Targeted Dimensions.

Parameter	**Value**	**Target**	**Model**	**Data**
p_1	244.69	Informality rate (output)	0.36	0.36
p_2	3.55	Informality rate (firms)	0.84	0.76
$\bar{\ell}$	0.17	Entrepreneurship rate	0.34	0.32
θ	0.14	Share informal firms ≤ 1	0.84	0.82
ρ	0.97	Share informal firms ≤ 2	1.00	0.95
σ_ν	0.06	Share formal firms ≤ 5	0.62	0.70

Table 6 displays how the model performs over some dimensions that are not targeted. These variables are the informality rates for self-employed and small firms (at most 5 employees), additional moments of the formal firm size distribution, including average firm size, and some public finance statistics.[13] Despite these variables are not targeted in the calibration, the model replicates them fairly well.

How does the model replicate the informality rates and the shares of small firms across

[13] As pointed out by Erosa et al. (2023), it seems more correct to compare sales tax revenues in the model to the sum of sales and income tax revenues in the data, since there is no income tax in the model.

Table 6: Data and Model Statistics over Untargeted Dimensions.

Moment	Model	Data
Informality rate among self-employed	0.90	0.91
Informality rate among firms ≤ 5	0.90	0.86
Share formal firms $6-10$	0.12	0.14
Share formal firms $11-20$	0.16	0.08
Share formal firms $21-50$	0.08	0.05
Share formal firms 50+	0.02	0.03
Average firm size	2.08	2.39
Sales tax/GDP	0.20	0.17
(Sales tax + Income tax)/GDP	0.20	0.24
Payroll tax/GDP	0.07	0.07

different size categories? Figure 2 depicts the shares of formal and informal firms for self-employed, firms with 2 employees (including the owner), and firms with 3-5 employees in the model (left columns) and in the data (right columns). While the model replicates well the informality rate for self-employed (90% v. 91% in the data), it overestimates it for firms with 2 employees (89% v. 66%) and it underestimates it for firms with 3-5 employees (0% v. 42%).

Figure 2: Informality Rates for Small Firms.

Source: Data from ECINF.

Further model implications. To conclude the model fit analysis, I present further statistics that do not have an exact analogous in the data but that replicate well-known patterns of Brazilian and other developing countries' economies.

Figure 3 shows the relationship between the number of employees and the probability of detection from the tax authority. The detection probability for a self-employed agent is between 1% and 7%[14], whereas it dramatically increases as employment goes up. In fact, the detection probability increases up to 73% for a firm with 2 employees and to 100% for a firm with 5 employees. This is consistent with the fact that about 95% of informal firms have two employees (including the owner) or less in the Brazilian data.

Figure 3: Probability of Detection for Different Levels of Employment.

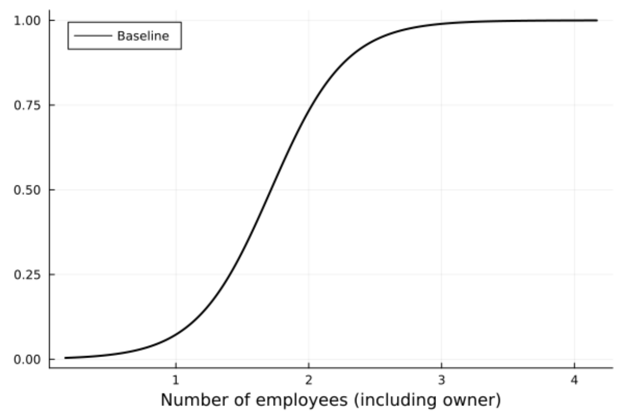


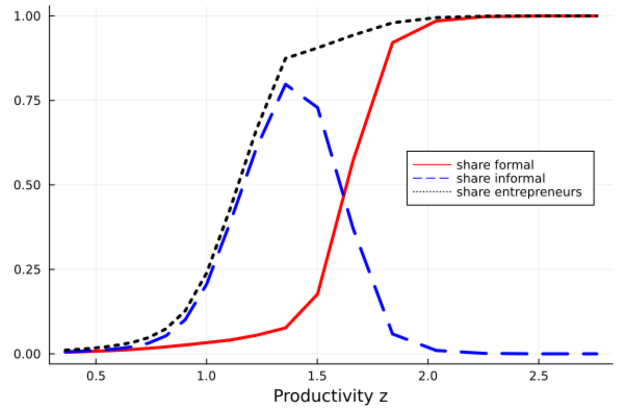
Figure 4 shows the share of formal (red solid line) and informal (blue dashed line) entrepreneurs over different levels of the productivity distribution. The dotted black line represents the overall share of entrepreneurs. For each share, the denominator is the total number of agents in the economy. The graph points out that informality is the preferred choice for entrepreneurs with low levels of productivity, but this preference is reversed as we move to the right of the distribution.

The graph is consistent with two patterns observed in the data: (i) more productive agents tend to be entrepreneurs rather than workers, and (ii) formal entrepreneurs are on average more productive than informal entrepreneurs. The latter pattern is observed in cross-country data[15] and in Brazilian data as well. In fact, Figure 5 depicts the productivity distributions of formal (red solid line) and informal (blue dashed line) firms with at most 5 employees in

[14] Notice that number of employees is a continuous variable in the model, therefore the detection probability of a self-employed agent is not constant. In the model, a self-employed agent is defined as such if $\bar{\ell} + n \leq 1$.

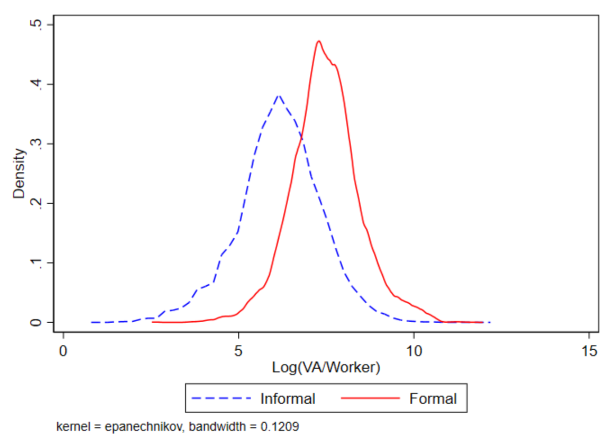
[15] For example, La Porta and Shleifer (2014) highlight the relationship between formality status and productivity across several countries using World Bank Enterprise Surveys data.

Figure 4: Share of Formal, Informal, and Entrepreneurs over the Productivity Distribution.



the ECINF data. Productivity is approximated as value added per worker.

Figure 5: Kernel Distribution of Productivity for Small (≤ 5) Formal and Informal Firms.

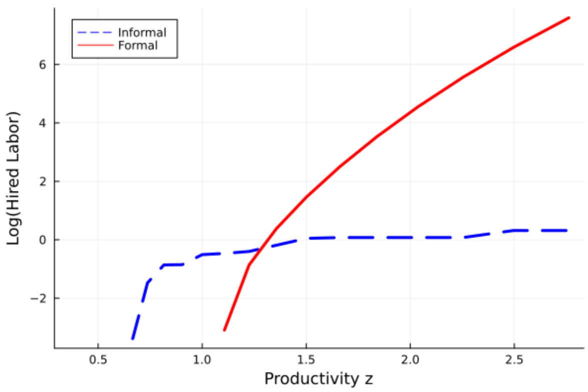


Productivity is measured as value added per worker (in logs). Source: Own calculations based on ECINF.

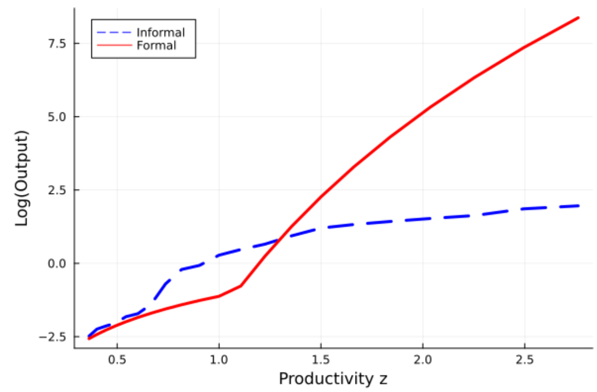
Finally, Figure 6 shows optimal labor and output policies for formal and informal entrepreneurs. In addition to the composition effect in productivity already described, informal entrepreneurs hire fewer employees compared to formal entrepreneurs with the same level of productivity (panel a). The reason for this behavior is to avoid detection from the tax authority. Consequently, the optimal choice of output is also lower for a given level of productivity (panel b).[16]

[16] This is not true for a few low productivity values. At these values, formal firms have to pay (distortionary) taxes, while informal firms do not and face a probability of detection that is close to 0. However, this effect tends to be quantitatively small in the aggregate, given the low mass of formal firms at these productivity levels.

Figure 6: Employment and Output for Different Productivity Levels.



(a) Log of employment over the productivity distribution for formal and informal.



(b) Log of output over the productivity distribution for formal and informal.

4 Quantitative Experiments

In this section, I perform some counterfactual experiments to assess the effects of tax enforcement on the variables of interest. The purpose of these exercises is to estimate the extent to which variations in tax enforcement, indirectly measured through variations in informality, can account for the observed differences in output, TFP, and average firm size. In the experiments, I vary the parameter p_1 of the detection probability function to match other countries' informality rates. I also adjust the sales tax rate τ_y to keep total revenues as close as possible to the baseline amount. All other parameters remain unchanged. Therefore, we can interpret the counterfactual economy as a version of the Brazilian economy characterized by a counterfactual tax enforcement.

4.1 Main Counterfactual

The main counterfactual matches the informality rate of the GDP-weighted average of the six largest Latin American economies, namely, Argentina, Chile, Colombia, Mexico, Peru, and Venezuela. The last column in Table 7 shows the average statistics for the counterfactual economy. The informality rate stands at around 30%, 6 percentage points lower than in Brazil. GDP per capita and TFP are about 32% and 6% higher than in Brazil. Finally, average firm size is 3.45, compared to 2.39 in the baseline.

In order to match the lower informality rate, the parameter p_1, which controls the inflection point of the detection probability function, goes down from a value of 245 to a value of 184. Figure 7 shows the probability of detection function as a function of number of employees in the baseline (black solid line) and in the counterfactual economy (gray dashed line). The probability of detection for a self-employed and for a firm with two employees are as high as 9% and 79%, respectively, compared to 7% and 73% in the baseline.

Figure 7: Probability of Detection for Different Levels of Employment, Baseline and Counterfactual.

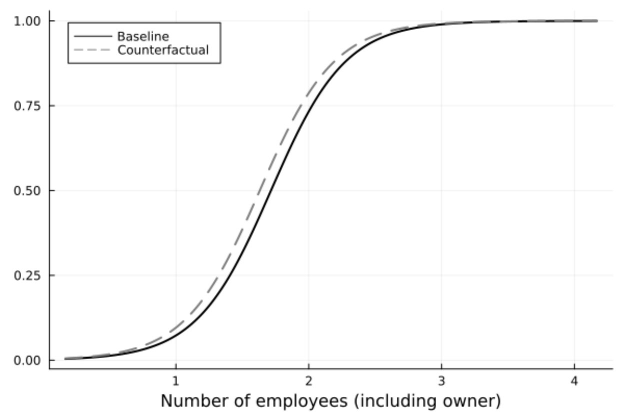


Table 7 points out that a stricter tax enforcement results in increased output, TFP, and average firm size. Compared to the baseline, output increases by 4.3 percentage points, which accounts for 13% of the measured difference in GDP per capita in 2003. TFP also goes up by about 0.5 percentage points, which is approximately 9% of the measured difference in TFP in 2003.[17] Average firm size increases from 2.08 in the baseline to 2.37 in the counterfactual. Bento and Restuccia (2021) estimate average firm size to be 2.39 in Brazil, while the figure for the GDP-weighted average of the six countries considered is 3.45. Therefore, stricter tax enforcement accounts for almost 30% of the estimated differences in average firm size. In this experiment, I change the sales tax rate to $\tau_y = 0.268$ ($\tau_y = 0.293$ in the baseline) to maintain total revenues close to the baseline amount. Therefore, the experiment suggests that stricter tax enforcement would not only reduce the level of informality but also allow for the reduction of taxes that are inefficiently high for the most productive agents in the

[17]GDP and TFP data are from Feenstra et al. (2015). In the model, TFP is measured according to the formula $TFP = \frac{Y}{(K^{\alpha}(L+s_e\bar{\ell})^{1-\alpha})^{\gamma}}$, where K. L, and Y denote aggregate capital, labor, and output, whereas s_e is the share of entrepreneurs in the population.

Table 7: Matching Counterfactual's Weighted Average Informality Rate - Counterfactual Moments. Share Accounted for by the Model in Brackets.

Variable	**Brazil**		**Counterfactual**	
	Model	**Data**	**Model**	**Data**
Informality rate	0.365	0.360	0.296	0.298
GDP per capita	1.000	1.000	1.043 [13.2%]	1.325
TFP	1.000	1.000	1.005 [8.8%]	1.062
Average firm size	2.080	2.390	2.374 [27.5%]	3.450

The counterfactual economy's data are computed as GDP-weighted averages of Argentina, Chile, Colombia, Mexico, Peru, and Venezuela. Source: own calculations based on Elgin et al. (2021) (informality), Feenstra et al. (2015) (GDP per capita and TFP), and Bento and Restuccia (2021) (average firm size). Data are for 2003 except for average firm size.

economy.[18]

4.2 Additional Experiments

This section describes additional quantitative experiments taking different countries or group of countries as counterfactuals. Besides considering single countries such as Argentina and Chile, I also perform an exercise that takes GDP-weighted average informality of a larger group of Latin American countries as counterfactual.[19] As for the previous exercise, the detection probability parameter p_1 is changed to match the correspondent level of informality for each country or group. I then compare model-implied GDP-weighted averages of GDP per capita, TFP, and average firm size with their measured counterpart.

Results of the counterfactual experiments are shown in Table 8. The top half displays measures of informality, GDP per capita, TFP, and average firm size for each country and group of countries. The bottom half presents the corresponding model-implied statistics. The percentages in brackets denote the shares of the measured differences in the variable of interest that are accounted for by the model.

[18]It has been documented that several developing economies are characterized by the coexistence of high informality rates and high tax rates (for the minority of agents and firms paying taxes).

[19]The group includes 17 countries: Argentina, Bolivia, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, Guatemala, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Uruguay, Venezuela.

Table 8: Effects of Tax Enforcement. Data and Model Statistics, Share Accounted for by the Model in Brackets.

Country	**Informality (1)**	**GDP pc (2)**	**TFP (3)**	**Average Size (4)**
Data				
Brazil (*Baseline*)	0.36	1.000	1.000	2.39
Group 1 (*Main*)	0.30	1.325	1.062	3.45
Group 2	0.31	1.248	1.051	3.42
Argentina	0.23	1.334	1.104	4.14
Chile	0.19	1.307	1.050	5.67
Mexico	0.29	1.567	1.187	2.83
Peru	0.55	0.633	0.675	2.45
Model				
Brazil (*Baseline*)	0.36	1.000	1.000	2.08
Group 1 (*Main*)	0.30	1.043[13.2%]	1.005[8.8%]	2.37[27.5%]
Group 2	0.31	1.034[13.9%]	1.004[7.9%]	2.30[21.0%]
Argentina	0.23	1.097[29.1%]	1.012[11.1%]	2.82[42.1%]
Chile	0.19	1.115[37.6%]	1.013[26.4%]	3.04[29.2%]
Mexico	0.29	1.047[8.3%]	1.006[3.3%]	2.40[71.5%]
Peru	0.55	0.919[22.1%]	0.975[7.8%]	1.69[-]

Source: Informality rates are from Elgin et al. (2021). GDP per capita and TFP are from Feenstra et al. (2015). Average firm size are from Bento and Restuccia (2021). Data are for 2003 except for average firm size.
Group 1: GDP-weighted average of Argentina, Chile, Colombia, Mexico, Peru, Venezuela. Group 2: GDP-weighted average of Argentina, Bolivia, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, Guatemala, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Uruguay, Venezuela.

Table 8 suggests that tax enforcement differences account for a relatively large fraction of the observed differences in average firm size, ranging from a minimum of 21.0% (Group 2) to a maximum of 71.5% (Mexico). On the other hand, the effects on GDP per capita and TFP account for a smaller fraction, ranging from a minimum of 8.3% and 3.3% (Mexico) to a maximum of 37.6% and 26.4% (Chile).

The low effect on TFP requires further discussion. It is important to remember that the only parameters that differ across experiments are the detection probability parameter p_1 and the sales tax rate τ_y, while all the other parameters have the same value as in the baseline Brazilian economy. Importantly, the invariant distribution of the productivity process is kept constant across experiments. Therefore, differences in aggregate productivity can only stem from heterogeneity in occupational choices and in capital and labor policy decisions. It can be argued that at least part of the observed differences in TFP is due to differences in the distribution of skills that cannot be accounted for by the model.

4.3 Economic Mechanisms

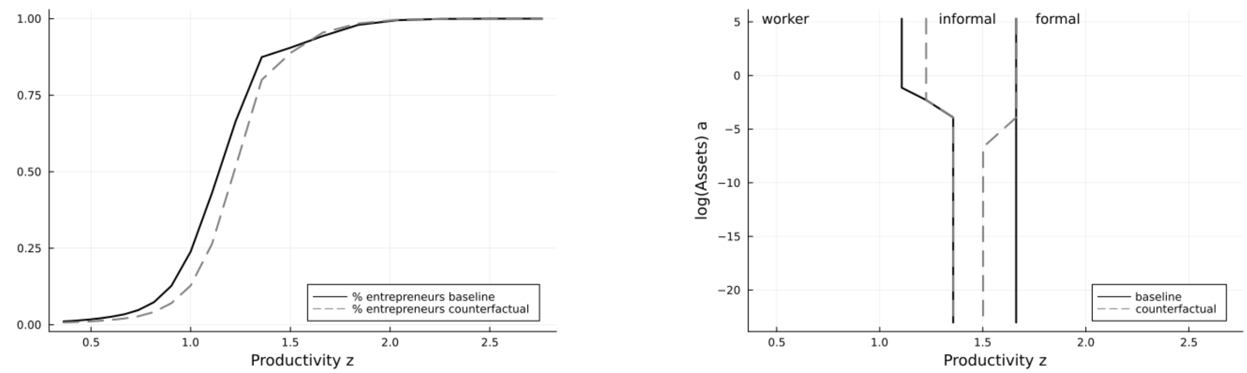
What are the underlying sources of the differences between the baseline and the counterfactual economies? To answer this question, it is useful to compare the occupational choices and the policy functions in the two economies. For illustration purposes, I will compare occupational and policy functions in the baseline and in the counterfactual where tax enforcement is such that it matches the Chilean informality rate (19%).[20] In the following graphs, the solid lines represent the values for the baseline economy, while the dashed lines represent the counterfactual values.

Figure 8 panel (a) shows that the share of entrepreneurs in the counterfactual economy is significantly lower for low values of productivity. Therefore, stricter tax enforcement leads to higher average entrepreneurial ability, inducing in turn a positive effect on average firm size given the positive relationship between firm-level productivity and size. Panel (b) displays occupational choices for different levels of productivity z (x-axis) and log-assets a (y-axis). The lines demarcate three areas, which correspond to combinations of productivity and assets where it is optimal to either be workers, informal entrepreneurs, or formal entrepreneurs.[21]
Notice that, for some values of productivity, agents with sufficient assets prefer to become informal entrepreneurs, while agents with low assets choose to be workers or formal entrepreneurs. This is due to the uncertainty in detection by the tax authority. In other words, for certain productivity values, the expected earnings are higher if agents decide to be informal entrepreneurs, but they only make this choice if they have enough assets. The graph shows not only that the share of entrepreneurs is lower in the counterfactual, but also that the formality rate among entrepreneurs is higher. This is caused by a combination of stricter detection probability and lower tax rates. Taken together, the two panels suggest that informal entrepreneurs reallocate either as workers or as entrepreneurs in the formal

[20]In this calibrated economy, $p_1 \approx 90$ and $\tau_y = 0.228$.

[21]In reality, the presence of occupational shocks leads to a probability distribution over occupational choices for each combination of productivity and assets. Therefore, the graph represents the choice with the highest probability.

Figure 8: Share of Entrepreneurs and Occupational Choices.



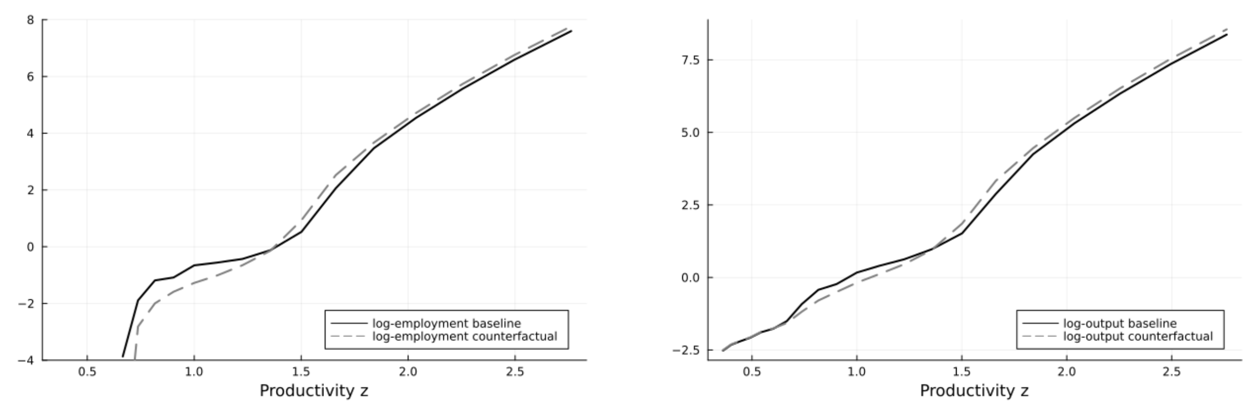
(a) Share of entrepreneurs (formal and informal) over the productivity distribution.

(b) Occupational choices for different levels of productivity z (x-axis) and log-assets a (y-axis).

sector.

The reallocation effect can also be observed in Figure 9, which depicts the average log-employment and log-output (conditional on being entrepreneurs) as a function of productivity in the two economies.

Figure 9: Employment and Output for Different Productivity Levels.



(a) Log of employment over the productivity distribution.

(b) Log of output over the productivity distribution.

Notice that employment and output are lower in the counterfactual economy for low values of productivity. There are two reasons for this pattern. First, for very low levels of productivity, formal firms tend to be smaller than informal firms. This is a consequence of the fact that small formal firms have to pay taxes, while small informal firms do not and face a detection probability which is close to 0. Second, as pointed out in Figure 7, for a given level of employment, informal firms face a higher detection probability in the counterfactual economy. Therefore, if they choose informality in the counterfactual economy, they might choose a lower scale to reduce the detection probability. However, these effects are quantitatively small in the aggregate and are more than compensated by higher values of labor and output for higher levels of productivity, as shown by the dashed line being on top of the solid line for both employment and output. The graphs point out to a reallocation of labor and output from low-productivity agents to high-productivity agents, which in turn leads to an increase in TFP and average firm size.

4.4 Decomposing Selection and Detection Channels

While in Section 4.3 I presented a qualitative analysis of the channels through which tax enforcement affects the outcomes of interest, in this section I quantitatively assess the contribution of the two main channels—selection and detection channels—to the overall estimated impact. To decompose the effects, I shut down one channel at a time, and then compare TFP, output, average firm size, and other statistics when only some decisions are allowed to change. The counterfactual considered is the GDP-weighted average of the six largest Latin American economies (*group 1*).

Table 9 displays the decomposition. Since the invariant distribution $\mu(z, a)$ is different in the counterfactual, I first fix the distribution as in the baseline in columns (2) to (4), while I adjust it to the counterfactual distribution in column (5). The selection effect in column (2) is obtained by maintaining the capital and labor decisions and the share of informality for every productivity level as in the baseline while allowing the share of entrepreneurs to change as in the counterfactual. Since the number of entrepreneurs goes down and production for the remaining entrepreneurs remains constant, output mechanically decreases. Despite an increase in average entrepreneurial ability, TFP does not seem to be affected by the selection channel. On the other hand, average firm size increases from 2.08 to 2.16.

Column (3) displays the contribution of the detection channel, which is obtained by maintaining the same share of entrepreneurs as in the baseline while allowing changes in the capital, labor, and formality status. The decrease in informality over the firm-level productivity distribution induces an overall increase in aggregate capital, labor, and output.

Moreover, the reallocation of resources towards more productive and larger firms implies an increase in TFP.

Table 9: Decomposition of Selection and Detection Channels.

	Baseline (1)	(2)	(3)	(4)	**Counterfactual** (5)
Fixed decisions	-	ℓ, k, form.	% entr.	-	-
Distribution $\mu(z,a)$	-	Baseline	Baseline	Baseline	Counterfactual
Channels	-	Selection	Detection	Sel.+Det.	Sel.+Det.+Distr.
Outcome					
Output	1.000	0.982	1.078	1.061	1.043
Capital	4.54	4.44	4.98	4.88	4.78
Labor	0.66	0.65	0.71	0.70	0.69
TFP	1.000	1.000	1.004	1.007	1.005
Average firm size	2.08	2.16	2.24	2.33	2.37
Entrepreneurship	0.34	0.32	0.34	0.32	0.31
Informality (Firms)	0.84	0.84	0.82	0.81	0.82
Informality (Output)	0.36	0.35	0.31	0.30	0.30

Column 1: baseline. Column 2: fixed capital and labor decisions, change in occupational choices allowed (*selection*). Column 3: fixed occupational choices, changes in capital and labor decisions allowed (*detection*). Column 4: sum of both effects keeping the same distribution $\mu(z,a)$ as in the baseline. Column 5: distribution adjusted (main counterfactual).

Column (4) shows the sum of the two effects while keeping the same invariant distribution as in the baseline. Two properties stand out. First, the rise in output due to the detection channel more than compensates the negative impact of the selection channel. Second, the interaction between the two channels seems to produce a positive effect on TFP and average firm size that is larger than the sum of the two channels taken in isolation.

Finally, in column (5), I allow the distribution to vary to obtain the counterfactual statistics already discussed in Table 7.

5 Conclusion

This paper develops a quantitative framework to assess the impact of tax enforcement on the firm size distribution and TFP. As tax enforcement becomes stricter, fewer agents choose to be entrepreneurs, and fewer entrepreneurs choose to operate in the informal sector. These mechanisms lead to a rise in average firm size and an improvement in allocative efficiency. Counterfactual analysis suggests that stricter tax enforcement in Brazil would result in sizable gains in average firm size, TFP, and GDP per capita. Considering several counterfactual experiments, the gains account for about 25-30% of the estimated differences in average firm size and for about 8-15% of the estimated differences in TFP and GDP per capita.

Therefore, this paper identifies tax enforcement as a relevant source of the observed cross-country differences in average firm size, TFP and GDP per capita.

References

Alvarez, J. and C. Ruane (2023). Informality and aggregate productivity: The case of Mexico. *Working Paper*.

Bachas, P., R. N. Fattal Jaef, and A. Jensen (2019). Size-dependent tax enforcement and compliance: Global evidence and aggregate implications. *Journal of Development Economics 140*, 203–222.

Bento, P. and D. Restuccia (2017). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics 9*(3), 267–303.

Bento, P. and D. Restuccia (2021). On average establishment size across sectors and countries. *Journal of Monetary Economics 117*, 220–242.

Buera, F. J., B. Moll, and Y. Shin (2013). Well-intended policies. *Review of economic dynamics 16*(1), 216–230.

Data Zoom (2023). Data Zoom: Simplifying access to brazilian microdata. Available at `https://www.econ.puc-rio.br/datazoom/english/index.html`.

Di Nola, A., G. Kocharkov, A. Scholl, and A.-M. Tkhir (2021). The aggregate consequences of tax evasion. *Review of Economic Dynamics 40*, 198–227.

Elgin, C., M. A. Kose, F. Ohnsorge, and S. Yu (2021). Understanding informality. *Working paper*.

Erosa, A., L. Fuster, and T. R. Martinez (2023). Public financing with financial frictions and underground economy. *Journal of Monetary Economics 135*, 20–36.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The next generation of the Penn World Table. *American Economic Review 105*(10), 3150–3182.

Franjo, L., N. Pouokam, and F. Turino (2022). Financial frictions and firm informality: A general equilibrium perspective. *Economic Journal 132*(645), 1790–1823.

Garicano, L., C. Lelarge, and J. van Reenen (2016). Firm size distortions and the productivity distribution: Evidence from France. *American Economic Review 106*(11), 3439–3479.

Guner, N., G. Ventura, and Y. Xu (2008). Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics 11*(4), 721–744.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Jounal of Economics 124*(4), 1403–1448.

La Porta, R. and A. Shleifer (2014). Informality and development. *Journal of Economic Perspectives 28*(3), 109–26.

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

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics 11*(4), 707–720.

Schneider, F., A. Buehn, and C. Montenegro (2010). New estimates for the shadow economies all over the world. *International Economic Journal 24*(4), 443–461.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence form Brazil. *American Economic Review 108*(8), 2015–2047.

A Appendix

A.1 Cross-Country Measures of Informality

To use measures of informality that are consistent across countries and over time, I rely on the database in Elgin et al. (2021), which provides estimates of informality rates for more than 160 countries over the period 1990-2019.

Elgin et al. (2021) introduce two time series who stem from different methods and assumptions. The first method, whose associated database I employ in the paper for cross-country comparisons, is based on a *dynamic general equilibrium* (DGE) model developed by Elgin and Öztunali (2012). In the model, a representative household decides how much labor to allocate to the formal and to the informal sector.

$$
\begin{aligned}
&\max_{\{C_t, X_t, N_{st}, N_{ft}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t U(C_t) \\
&\text{s.t.} \quad C_t + X_t = (1-\tau_t)\theta_{F_t} K_t^{\alpha} N_{F_t}^{1-\alpha} + \theta_{S_t} N_{S_t}^{\gamma}, \\
&\qquad\quad K_{t+1} = X_t + (1-\delta)K_t, \\
&\qquad\quad N_{S_t} + N_{F_t} = H_t,
\end{aligned}
\tag{5}
$$

where N_{F_t}, N_{S_t} are the amounts of hours the household devotes to the formal and informal sector, while C_t, X_t, and K_t denote consumption, investment, and capital, respectively. The various parameters and variables are either externally set according to conventional values, or retrieved by combining national accounts data and the first order conditions from (5). An additional assumption is needed to derive θ_{S_t}, the productivity growth rate in the informal sector: it is assumed to be the average between the productivity growth in the formal sector θ_{F_t}, and the growth rate of capital. Notice that this procedure retrieves the growth rate of the informal sector, not its absolute value. Therefore, for each country the share of informal output over GDP in 2007 is set to be equal to the shadow economy size as estimated in Schneider et al. (2010) for the same year.

The *Multiple Indicators Multiple Causes* (MIMIC) in Schneider et al. (2010) constitutes the source of the second series in Elgin et al. (2021). The method consists of a simultaneous specification of a factor (measurement) model and a structural model. The main idea is to retrieve an unobserved variable using some structural equations and the sample covariances between observed variables.

The structural model can be written as

$$\eta = \gamma' X + \xi,$$

where η is the unobserved informal economy over GDP ratio, and X is a vector of explanatory variables (causes) such as the share of direct taxes over overall taxes, government expenditures over GDP, and fiscal freedom, business freedom, and regulatory indices.

The measurement model can be written as

$$y = \lambda \eta + \epsilon,$$

where y is a vector of indicator variables such as currency (M0 or M1), GDP per capita growth, and labor force participation. By combining the two equations, one can derive a relationship between causes X and indicators y, as well as the related covariance matrix. To obtain the coefficients needed to infer the size of the informal economy η, the authors use maximum likelihood in order to minimize the distance between theoretical and sample covariances. This second method also allows for computing the relative size of the shadow economy, not the absolute size. Therefore, estimates based on a currency approach for the year 2000 are used as base year.

A.2 Cross-Country Relationship between Informality and TFP

Table 10 shows the world averages of informal share of GDP and TFP from 1990 to 2018. Using the DGE based measure of informality as a benchmark (column 1), the table highlights that the informal share of GDP has decreased by about 8 percentage points from 34.8 to 26.8 in the period considered.[22] In the same period, TFP has increased by about 6 percentage points on average (column 4). Column (3) displays the average estimate of TFP

[22] Informality has also decreased according to the MIMIC measure of informality, but by about 3 percentage points (column 2).

as a percentage of US TFP. This is a measure used for comparisons of TFP across countries at a given point in time.

Figure 10 depicts the cross-country correlation between informality and TFP. The left panel shows the DGE-based measure, while the right panel displays the MIMIC-based measure. TFP is measured as a percentage of US TFP. The two panels point out a negative correlation between TFP on the one hand and both measures of informality on the other hand.

Table 10: Informality and TFP. Simple Averages.

Year	% Informal GDP (DGE) (1)	% Informal GDP (MIMIC) (2)	TFP (US=100) (3)	TFP (2017=100) (4)
1990	34.8	.	72.4	94.4
1991	34.7	.	71.0	93.1
1992	34.3	.	69.9	93.8
1993	34.3	34.7	68.7	93.6
1994	34.1	34.6	65.4	88.8
1995	34.0	34.5	65.5	88.8
1996	33.8	34.3	65.7	89.5
1997	33.6	34.2	65.2	90.4
1998	33.4	34.2	63.5	90.3
1999	33.2	34.2	63.3	90.3
2000	33.1	34.0	65.3	91.1
2001	32.9	34.0	65.0	91.6
2002	32.7	34.1	65.3	92.6
2003	32.5	34.0	64.8	93.7
2004	32.3	33.7	65.1	96.2
2005	32.2	33.6	67.2	97.5
2006	31.9	33.2	68.0	99.2
2007	31.6	32.9	68.4	100.4
2008	31.3	32.8	68.0	99.9
2009	30.9	33.4	64.8	97.8
2010	30.7	33.0	64.7	99.3
2011	30.4	32.9	66.8	100.1
2012	30.0	32.8	67.4	100.4
2013	29.7	32.7	66.9	100.4
2014	29.4	32.6	65.9	100.3
2015	29.1	32.5	63.8	99.7
2016	28.9	32.4	63.7	99.6
2017	28.7	32.1	64.4	100.0
2018	26.8	31.9	64.0	99.9

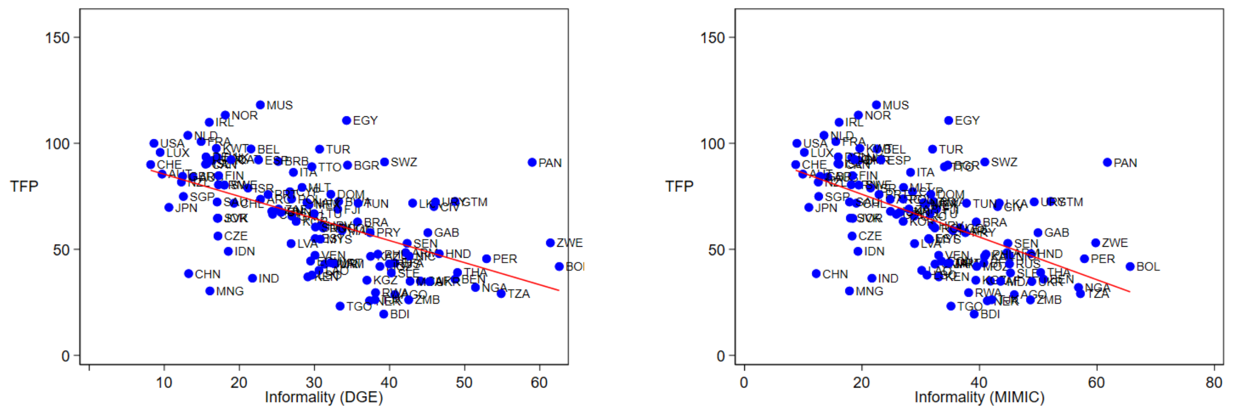
World averages. Sources: Informality data are from Elgin et al. (2021). TFP data are from Feenstra et al. (2015).

To further assess the relationship between informality and TFP, I perform a series of cross-country regressions. First, I run an OLS regression according to the following specification:

$$TFP_{it} = \alpha_0 + \alpha_1 INF_{it} + \gamma GDP_{it} + Year_t + \epsilon_{it},$$

Figure 10: Correlation between Informality and TFP across Countries.

(a) Informality (DGE) and TFP (US=100). (b) Informality (MIMIC) and TFP (US=100).



Data sources: Elgin et al. (2021) for informality, Feenstra et al. (2015) for TFP. Each data point represents the country's averages of informality and TFP over the period between 1990 and 2018.

where TFP_{it} is TFP in country i in year t (relative to US TFP in year t), INF_{it} is the share of informal output, GDP_{it} is GDP per capita, $Year_t$ denotes year fixed effects, and ϵ_{it} is the error term. α_1 is the coefficient of interest in the relationship between informality and TFP. Since the two series of informality are highly correlated, I adopt the DGE-based measure, which is the one used in the quantitative experiments in Section 4.

Second, I exploit the time series dimension of the data to perform a panel regression:

$$TFP_{it} = \alpha_0 + \alpha_1 INF_{it} + \gamma GDP_{it} + c_i + \epsilon_{it},$$

where c_i denotes country-specific fixed effects. Since this type of panel regression captures the effect within a country, the correct measure of TFP to use is the one relative to a base year (2017) for that country, that is, the series whose cross-country averages are shown in column (4) of Table 10.

Table 11 shows the results of the regressions described above. Column (1) displays the coefficient of the OLS specification without GDP per capita, while column (2) includes it. The inclusion of GDP per capita reduces the magnitude of the coefficient of interest, which remains negative and statistically significant. A coefficient of -0.26 can be interpreted in the following way: a decrease in informality by one percentage point is associated with an

Table 11: Relationship between TFP and Informality.

	(1) OLS TFP (US=100)	(2) OLS TFP (US=100)	(3) Panel TFP (2017=1)	(4) Panel TFP (2007=1)
Informality (DGE)	-1.026***	-0.261***	-0.879***	-0.852**
	(0.0313)	(0.0344)	(0.321)	(0.339)
GDP per capita (th)		0.828***		0.0271
		(0.0237)		(0.227)
Year fixed effects	YES	YES	NO	NO
Country fixed effects	NO	NO	YES	YES
Observations	3198	3198	3198	3198

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

increase in TFP (relative to US TFP) by 0.26 percentage points. Column (3) and column (4) display the results of the panel regressions excluding and including GDP per capita as control, respectively. The coefficients on informality are negative and statistically significant. In this case, a coefficient of -0.85 means that a reduction in informality by one percentage point is associated with an increase in TFP by 0.85 percentage points.

A.3 Simple Version of the Model

It is useful to consider a simplified static version of the model to highlight the main mechanisms that carry over into the dynamic model presented in Section 2.[23] In this version, owner's labor $\bar{\ell}$ is set equal to 0. Moreover, the probability of detection is a discrete function: it jumps from 0 to 1 if the number of employees is larger than a threshold $\bar{n}$, that is,

$$p(n) = \begin{cases} 0 & \text{if } n \leq \bar{n}, \\ 1 & \text{if } n > \bar{n}. \end{cases}$$

This last assumption eliminates the uncertainty regarding detection from the tax authority. As a consequence, agents choose the occupation that gives them the highest earnings. As long as $\bar{n}$ is not too low and τ_y not too high, there will be three productivity thresholds z_1,

[23]This version is close to the model in Leal Ordóñez (2014).

z_c, and z_2 such that:

(a) Each agent with $z \leq z_1$ is a worker.

(b) Each agent with $z_1 < z \leq z_c$ is an unconstrained ($n < \bar{n}$) informal entrepreneur.

(c) Each agent with $z_c < z \leq z_2$ is a constrained ($n = \bar{n}$) informal entrepreneur.

(d) Each agent with $z > z_2$ is a formal entrepreneur.

The intuition behind this result is the following. First, notice that labor earnings w do not depend on productivity z. On the other hand, informal entrepreneur's profits $\pi_I(z)$ are increasing in z and equal to

$$\pi_I(z) = \begin{cases} \gamma^{\frac{\gamma}{1-\gamma}}(1-\gamma)z^{\frac{1}{1-\gamma}}\left(\frac{\alpha}{r+\delta}\right)^{\frac{\gamma\alpha}{1-\gamma}}\left(\frac{1-\alpha}{w+\theta}\right)^{\frac{\gamma(1-\alpha)}{1-\gamma}} & \text{if } z_1 < z \leq z_c, \\ \gamma^{\frac{\gamma\alpha}{1-\gamma\alpha}}(1-\gamma\alpha)z^{\frac{1}{1-\gamma\alpha}}\left(\frac{\alpha}{r+\delta}\right)^{\frac{\gamma\alpha}{1-\gamma\alpha}}\bar{n}^{\frac{\gamma(1-\alpha)}{1-\gamma\alpha}} - (w+\theta)\bar{n} & \text{if } z_c < z \leq z_2. \end{cases}$$

Therefore, z_1 is the productivity value such that $\pi_I(z_1) = w$.
Formal entrepreneur's profits $\pi_F(z)$ are also increasing in z and equal to

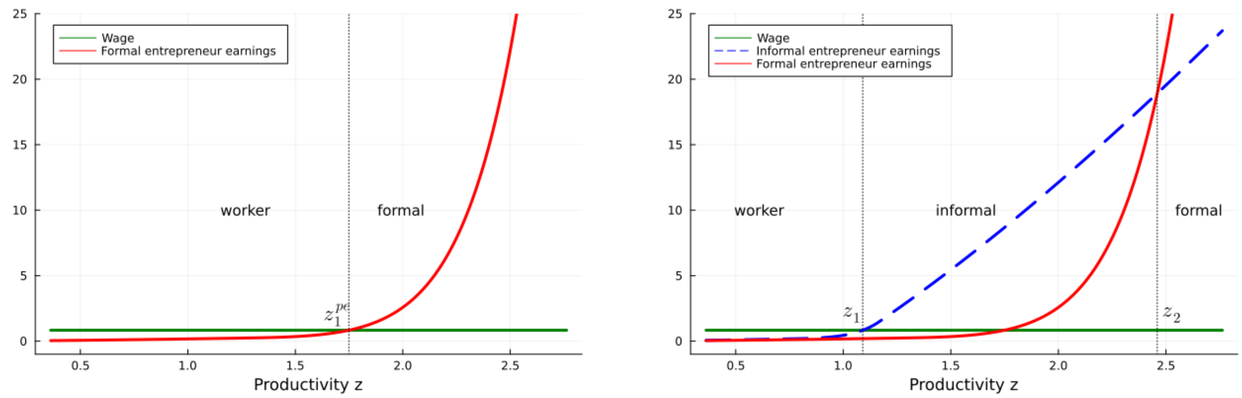
$$\pi_F(z) = \gamma^{\frac{\gamma}{1-\gamma}}(1-\gamma)((1-\tau_y)z)^{\frac{1}{1-\gamma}}\left(\frac{\alpha}{r+\delta}\right)^{\frac{\gamma\alpha}{1-\gamma}}\left(\frac{1-\alpha}{w(1+\tau_n)+\theta}\right)^{\frac{\gamma(1-\alpha)}{1-\gamma}}.$$

The profits equations point out that $\pi_F(z) < \pi_I(z)$ for each value of z such that $z_1 < z \leq z_c$, because informal entrepreneurs do not pay taxes. However, since $\frac{1}{1-\gamma} > \frac{1}{1-\gamma\alpha}$, as productivity z grows, the profit function for formal entrepreneurs increases at a faster rate than the profit function for informal entrepreneurs with $n_I(z) = \bar{n}$. The economic intuition is that formal entrepreneurs can expand their scale and choose the optimal combination of capital and labor. Therefore, there exists z_2 such that $\pi_I(z_2) < \pi_F(z_2)$.
I now perform a comparative statics exercise between two economies that differ in their detection probability parameter $\bar{n}$. First, I consider a *perfect enforcement* economy characterized by the absence of an informal sector since $\bar{n} = 0$. In this economy, there is a productivity threshold z_1^{PE} such that all agents with $z \leq z_1^{PE}$ are workers and all agents with $z > z_1^{PE}$ are entrepreneurs. Figure 11 panel (a) depicts labor earnings (green line) and profits (red line) over the productivity distribution. The threshold z_1^{PE} is the value at which entrepreneurial

profits are equal to the wage rate w. All the agents above the threshold choose to be (formal) entrepreneurs. In an *imperfect enforcement* economy with $\bar{n} > 0$ high enough, however, an equilibrium is characterized by two productivity thresholds. Figure 11 panel (b) depicts labor earnings (green solid line), profits for formal (red solid line) and informal entrepreneurs (blue dashed line) in the imperfect enforcement economy. The threshold z_2 is the threshold above which agents decide to be formal entrepreneurs.

Figure 11: Earnings and Occupational Choices for Different Productivity Levels.

(a) Perfect Enforcement Economy. (b) Imperfect Enforcement Economy.

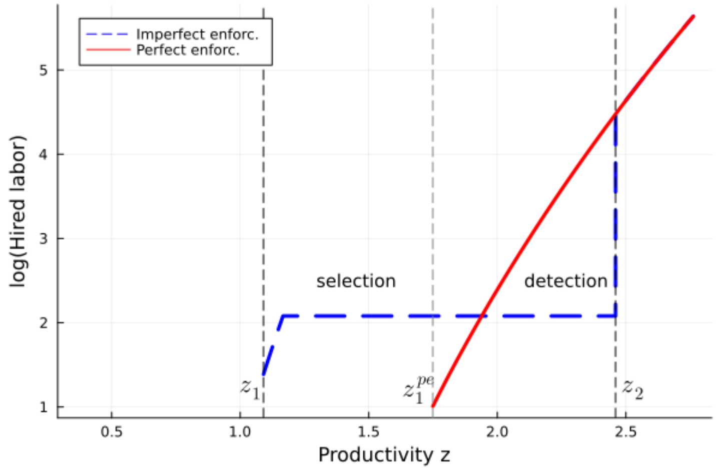


Labor earnings (green line), formal entrepreneurs profits (red solid line) and informal entrepreneurs profits (blue dashed line).

How is labor allocated among agents in the two economies? Figure 12 plots the log of labor hired by each agent in the perfect enforcement economy (red solid line) and in the economy with informality (blue dashed line). The graph highlights the two main mechanisms. First, in the economy with informality, more agents decide to become entrepreneurs, since the possibility of not paying taxes increases the value of entrepreneurship. Given the positive relationship between firm-level productivity and size, this *selection channel* leads to a decrease in both average firm size and average productivity.[24] Second, compared to the perfect enforcement economy, some entrepreneurs operate at lower scale and choose to be small enough not to pay taxes (*detection channel*). This induces a drop in average firm size, aggregate labor demand, and output.

[24]However, the selection channel partly offsets the negative effect on output stemming from the detection channel.

Figure 12: Hired Labor over the Productivity Distribution in the Two Economies.



Log of hired labor over productivity under perfect enforcement (red solid line) and imperfect enforcement (blue dashed line). The wage rate w is kept constant in the two economies (partial equilibrium).

A.4 Solution Method Algorithm

This appendix describes the algorithm used to get the value and policy functions introduced in Section 2. The recursive problem of a worker with productivity z and asset a is

$$V_W(z, a) = \max_{c,a'} \; u(c) + \beta \mathbb{E} V(z', a')$$
$$s.t. \quad c + a' = (1+r)a + w, \quad a' \geq 0,$$

where V is the value function of an agent before the employment decision is made.
First order conditions imply

$$u'(c) = \beta \mathbb{E} V_a(z', a'). \tag{B.1}$$

I can get an update of the value function V_W by using the endogenous grid method: (i) start with an exogenous asset grid over a', (ii) invert the Euler equation (B.1) and get an endogenous consumption grid, (iii) use the budget constraint to get an endogenous asset grid a_e, (iv) interpolate over a' values to get the asset policy function a_p and the expected value function $\mathbb{E}V(z', a_p)$.
However, because of the discrete employment decision, V is in general not concave and the dynamic programming is non-convex. The Euler equation (B.1) is therefore necessary but

not sufficient. As a consequence, the endogenous asset grid a_e might not be monotone, and there might be regions in which for one value of a_e there are multiple values of a'. To solve this issue, an additional step is necessary compared to the standard endogenous grid method. This step is called *upper envelope*, which consists in computing the value function in the region where there are multiple a' for a given a_e, pick the (a_e, a') pair that returns the highest value function, and discard the other pairs (Fella (2014), Iskhakov et al. (2017)). I can get $V_F(z,a)$ and $V_I(z,a)$ in a similar fashion.

Now that I have $V_o(z,a)$ for $o = \{W, F, I\}$, and since the taste shock is iid EV1, I can compute $V(z,a)$ using the logsum formula

$$V(z,a) = \sigma log\Bigg(\sum_{o=\{W,F,I\}} exp\bigg(\frac{V_o(z,a)}{\sigma}\bigg)\Bigg). \tag{B.2}$$

The algorithm can be summarized through the following steps:

1. Start with a guess $V(z,a')$ over an exogenous grid on z and a'.
2. Compute $\beta \mathbb{E} V_a(z', a')$ (using approximate derivatives in the first iteration, or the envelope condition in subsequent iterations).
3. Get $a_e(z,a)$ and $c_e(z,a)$ using the worker's first order conditions and budget constraint.
4. Interpolate the asset policy function, which is currently on the endogenous grid a_e, on the exogenous asset grid a' and get the asset policy a_p. Interpolate the next period expected value function on the optimal asset policy as well and get $\beta \mathbb{E} V(z', a_p)$.
5. *Upper envelope step:* For a given a_e, there might be multiple a' and, therefore, multiple a_p and $V_W(z,a')$. Keep the maximum $V_W(z,a')$ and the related asset/consumption policies, discard the others.
6. Obtain $V_F(z,a')$ and $V_I(z,a')$ similarly to steps 3-5.
7. Use the logsum formula (B.2) to get $V(z,a')$.
8. Check convergence. If convergence is not satisfied, go back to step 2 using the updated $V(z,a')$.

A.5 Brazilian Data

ECINF.[25] *Pesquisa de Economia Informal Urbana* (ECINF) is a survey conducted by *Instituto Brasileiro de Geografia e Estatística* (IBGE), the Brazilian Bureau of Statistics. It was conducted in 1997 and 2003 to collect information about the informal sector. The survey is nationwide representative for small non-agricultural businesses with a maximum of 5 employees. Owners are classified as informal if they do not possess a tax identification number (*Cadastro Nacional de Pessoa Juridica*, CNPJ).[26] By matching owners and businesses with employees, it is possible to obtain the number of employees for each business. Additional information such as sector and financial variables is also available.

From the original dataset, which is publicly available, I exclude the following observations:

1. Owners who operate in the agricultural and construction sectors.
2. Owners who lack a facility exclusively dedicated to the business outside their house.
3. Owners who have another job.
4. Businesses who have no owners or more than 4 owners.
5. Businesses who have more than 7 employees.

2 and 3 directly address the concern that the data might measure home production rather than entrepreneurship (as pointed out by Erosa et al. (2023)).

The final sample from which I compute moments used in the calibration contains about 30,000 firms. Table 12 summarizes sector composition and size distribution by formality status.

PNAD. I use data from the 2003 *Pesquisa Nacional por Amostra de Domicílios* (PNAD), a repeated cross-section representative at the national level, to compute the rate of entrepreneurship. Individuals who are either self-employed or employers are classified as entrepreneurs.

[25] Section A.5 has strongly benefited from the Data Zoom (2023) project and the Ulyssea (2018) replication package.

[26] Ulyssea (2018) points out that strict confidentiality clauses and IBGE's reputation induce respondents to report fairly accurately.

Table 12: ECINF. Sector and Size Statistics.

	Formal Shares	Informal Shares	Total Shares
Industry			
Manufacturing	8.69	11.27	10.82
Retail	50.50	37.70	39.93
Services	40.81	51.03	49.25
Number of Employees			
1	28.95	79.30	70.53
2	27.39	14.14	16.45
3	16.35	3.68	5.89
4	11.71	1.72	3.46
5	8.65	0.79	2.16
6	4.49	0.30	1.03
7	2.47	0.07	0.49
Observations (#)	5,261	24,924	30,185

Source: Own calculation from ECINF (2003). Note that the shares reflect the raw unweighted sample.

Appendices References

Elgin, C., M. A. Kose, F. Ohnsorge, and S. Yu (2021). Understanding informality. *Working paper*.

Elgin, C. and O. Öztunali (2012). Shadow economies around the world: Model based estimates. *Working Papers 2012/05. Bogazici University.*.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The next generation of the Penn World Table. *American Economic Review 105*(10), 3150–3182.

Fella, G. (2014). A generalized endogenous grid method for non-smooth and non-concave problems. *Review of Economic Dynamics 17*(2), 329–344.

Iskhakov, F., T. H. Jørgensen, J. Rust, and B. Schjerning (2017). The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks. *Quantitative Economics 8*(2), 317–365.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence form Brazil. *American Economic Review 108*(8), 2015–2047.