Market Power with Formal and Informal Firms*

Tomas Martinez[†] Cezar Santos[‡]

February 26th, 2026

Abstract

Market power among firms is a common feature in many economies, including in developing countries where a significant share of firms operate informally. This paper examines the interaction between market power and informality, and how this interaction shapes the impact of government policies. We develop a general equilibrium model with firm heterogeneity and oligopolistic output markets, in which firms endogenously decide whether to operate formally or informally. Informal firms evade taxes but incur enforcement costs. We calibrate the model using comprehensive administrative data from Brazil and perform several counterfactuals.

Keywords: Market Power, Informality, Markups, Market Structure, Taxes.
JEL codes: D4, D5, E1, L1.

*We have benefited from helpful comments from Jan Eeckhout. The views expressed in this article are those of the authors and do not necessarily represent those of the Inter-American Development Bank. Tomas gratefully acknowledges financial support from FAPESP grant 2024/20171-9 and from the Becker Friedman Institute at the University of Chicago.

[†]Insper. Email: tomas.martinez@insper.edu.br
[‡]Inter-American Development Bank & CEPR. Email: cezarsantos.econ@gmail.com

1 Introduction

There is substantial evidence of imperfect competition and market power across different regions of the world, including in developing countries (De Loecker and Eeckhout, 2021). The developing world is also characterized by a large informal sector (Ulyssea, 2020). Informal firms tend to be small, which can result in a large market share for formal firms, enabling them to exercise market power by charging higher markups and producing less output. Consequently, the interaction between informality and market power can influence the effects of government tax policies and the aggregate economy as a whole. This paper investigates these interactions by developing a quantitative general equilibrium model, calibrated using rich administrative data from Brazil.

We build a macroeconomic general equilibrium model featuring heterogeneous firms with market power, where firms choose to operate either formally or informally. Following an oligopolistic market structure, the final good is produced from inputs across a continuum of markets, each containing a finite number of firms. Firms decide whether to formalize based on their productivity and fixed costs. Formal firms pay taxes, while informal firms avoid taxes but face enforcement costs, which increase with firm size. The model captures the strategic interaction between firms, where they compete within markets (via Cournot competition) but take prices as given across markets.

Households derive utility from consumption and supply labor. They own the firms and receive lump-sum transfers from the government. The firm's problem is solved backward, first determining the production decision and then deciding whether to formalize. The formalization decision depends on profits, which are influenced by market power, taxes, and the costs of informality. The solution of the model yields equilibrium conditions for markups, market shares, and the distribution of firms across formal and informal sectors. Market-level and aggregate outcomes are affected by the interaction between market power and informality.

The model is calibrated using Brazilian data from several sources, including administrative datasets like PIA (industrial firms), RAIS (formal labor contracts), ECINF (informal sector data), and PNAD (household survey). The calibration focuses on matching key moments related to the formal and informal sectors, such as markups, firm size, and sales distributions. Parameters related to taxes, elasticities, and productivity dispersion are either externally calibrated based on existing literature and direct data counterparts or internally estimated by matching simulated moments to observed data.

The calibrated model is used to quantify the aggregate interaction between informality and market power and to assess the macroeconomic consequences of formalization policies. The key insight from the quantitative analysis is that informality has modest direct effects on markup dispersion, but large general-equilibrium effects through firm reallocation and competitive responses.

We begin by decomposing the aggregate cost of market power into two components: the level of markups and their dispersion. In the baseline economy, the aggregate markup is 1.238 and markup dispersion generates a misallocation wedge of 3.1 percent. Eliminating the level distortion raises output by 4.5 percent, while removing dispersion increases output by 3.1 percent. Shutting down both distortions simultaneously increases output by 7.8 percent. Thus, both average markups and their dispersion contribute meaningfully to aggregate losses.

We then examine the role of informality. In partial equilibrium, eliminating informality has only a small effect on markup dispersion. Although informal firms are numerous, they account for a limited share of production, and changes in their market shares largely offset across productivity regions. As a result, the aggregate markup and the misallocation wedge respond only modestly when informality is shut down holding other variables fixed.

In general equilibrium, however, the effects are substantially larger. Reducing informality by half increases output by 10.5 percent, while eliminating it entirely raises output by 21.3 percent. These gains arise from two complementary forces. First, firms that transition from informal to formal status expand significantly. Second, incumbent formal firms respond competitively to this expansion, increasing production despite modest increases in real wages. The largest output gains therefore reflect not only direct reallocation effects, but also the amplification generated by strategic interaction among firms.

Importantly, even in economies with lower informality, market power continues to generate sizable distortions. Setting aggregate markups to unity or eliminating dispersion still yields non-trivial output gains, underscoring that formalization alone does not eliminate inefficiencies arising from imperfect competition.

Our paper contributes to two strands of the literature. While previous studies, such as De Loecker et al. (2020) and Edmond et al. (2023), focus on documenting the rise of market power and its macroeconomic implications in developed economies like the United States, they overlook the role of informality, which is an important factor in many developing countries. Unlike these studies, our paper explicitly incorporates market power and informality together, enabling us to study how these factors jointly influence aggregate outcomes and government policy in a developing country context. Boar and Midrigan (2023) study the optimal policy in the presence of market power in a Mirleesian world. While we do not tackle optimal policies, we add by investigating the effects of tax policies in a context with both market power and informality.

Our approach also differs from studies like Ulyssea (2018), Meghir et al. (2015) and Erosa et al. (2023), which explore the implications of informality but do so without directly addressing the role of market power. Ulyssea (2018) introduces a model where firms exploit different margins of informality, and Erosa et al. (2023) examines the positive role informality can play under financial constraints. However, neither paper considers how market power might amplify or mitigate the effects of informality on productivity and welfare. By integrating these two dimensions—market power and informality—our paper provides a comprehensive framework for analyzing the broader economic implications, aligning more closely with De Loecker et al. (2022), which considers both technology and competition as drivers of market power, though still within a developed economy context. Our quantitative analysis, which examines the effects of tax and enforcement policies on informality and market power, further differentiates our study by offering concrete policy insights for developing economies, a perspective less explored in the existing literature.

The rest of this paper is organized as follows. Section 2 describes our model. Section 3 lays out the calibration procedure to discipline the model parameters. Section 4 discusses the quantitative results. Section 5 concludes.

2 Model

This section presents a model with heterogeneous firms that have product market power and decide whether to operate in the formal or informal sector. The model follows the oligopolistic market structure of Edmond et al. (2015) and De Loecker et al. (2022), where the final good is produced using inputs from a continuum of markets $s \in [0, 1]$, and each market contains a finite and exogenous number of firms, M_s. A firm, indexed by i, is randomly allocated in market s and produces a good indexed by is using capital and labor. Time is discrete and we focus on the stationary equilibrium.

2.1 Households

There is a representative household with unity mass that derives utility from the consumption of the final good and supplies labor elastically to the market. The household has a time-separable utility function and discounts future utility at a rate $\beta \in (0, 1)$. The household owns all the firms and receives all the aggregate tax revenue as lump-sum transfers. The household's maximization problem is given by:

$$
\begin{aligned}
\max_{\{C_t, N_t\}_{t=0}^{\infty}} \quad & \sum_{t=0}^{\infty} \beta^t u\left(C_t - \hat{\nu}^{-\frac{1}{\nu}} \frac{N_t^{1+1/\nu}}{1+1/\nu}\right) \\
\text{s.t.} \quad & P_t C_t + K_{t+1} \leq (1 + R_t - \delta) K_t + W_t N_t + T_t + \Pi_t \\
& K_0 > 0,
\end{aligned}
$$

where $u(.)$ is increasing and concave, C_t is aggregate consumption of the final good, N_t is labor supply, K_t is physical capital, P_t is the price index, W_t is the nominal wage, R_t is the rental rate of capital, T_t is aggregate tax revenue, and Π_t is the aggregate profit from all firms. We assume GHH preferences to abstract from wealth effects on labor supply.[1] The solution to the problem yields the standard Euler equation and the household's labor supply decision.

2.2 Producers

Final-good producers. Let Y be the output of the final good.[2] As in Atkeson and Burstein (2008), production is carried through a nested Constant Elasticity of Substitution (CES) aggregator. In the outer layer, the final good firm aggregates market-level intermediate inputs, y_s, from a continuum of markets:

$$
Y = \left(\int_0^1 y_s^{\frac{\theta-1}{\theta}} ds\right)^{\frac{\theta}{\theta-1}},
$$

where $\theta > 1$ is the elasticity of substitution across markets. The solution of the final good producer implies the following demand function: $y_s = (p_s/P)^{-\theta} Y$. The price index is:

$$
P = \left(\int_0^1 p_s^{1-\theta} ds\right)^{\frac{1}{1-\theta}} \tag{1}
$$

In the inner layer of the CES aggregator, the market-level output is produced using a *discrete* and market-specific number of M_s intermediate inputs:

$$
y_s = \left(M_s^{-\frac{1}{\gamma}} \sum_{i=1}^{M_s} y_{is}^{\frac{\gamma-1}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}}.
$$

[1] Changes in the market structure and informality affect aggregate taxes, profits and capital income, leading to wealth effects on labor supply. We abstract from these to focus solely on the direct effects of market power on the economy.

[2] Since the firm's problem is static, we omit time subscripts t to ease notation.

Again, the demand for a intermediate input, i, is $y_{is} = (p_{is}/p_s)^{-\gamma} y_s/M_s$, while the market-level price index is:

$$p_s = \left(M_s^{-1} \sum_{i=1}^{M_s} p_{is}^{1-\gamma} \right)^{\frac{1}{1-\gamma}}, \tag{2}$$

where γ controls the within-market elasticity of substitution. We assume that $\gamma > \theta > 1$. That is, the elasticity of demand within a particular market is higher than across markets, meaning intermediate goods are more substitutable within markets (e.g., different brands of coffee) than across markets (e.g., coffee and cars). Furthermore, we assume intermediate firms have market power and, given the finite number of firms within the market, engage in *Cournot* competition. Each firm chooses its optimal quantity y_{is}^* taking as given the optimal quantities decided by its competitor within a market. As it will become clear later, this does not mean that the elasticity of substitution across markets is not important. In the model, markets are just limits on the strategic behavior of firms: they consider the optimal strategy of their competitors within markets but act as price takers with respect to the firms in other markets.

Technology. Firms can choose to operate formally or informally and are ex-ante heterogeneous in two dimensions. First, firms differ in their productivity level, z_{is}. Second, they must pay a fixed cost if they operate in the formal sector. To produce the intermediate input, y_{is}, firms combine capital, k_{is}, and labor, n_{is}, according to the following constant returns to scale technology:

$$y_{is} = z_{is} k_{is}^{\alpha} n_{is}^{1-\alpha}.$$

After observing their productivity, firms self-select into the sector that offers the highest profit. We solve the firm's problem backward, first characterizing the production decision and then the sector selection. Below, we discuss the profit maximization problem in each sector.

Formal firms. Formal firms are subject to sales and labor taxes and must pay a fixed operating cost. The total cost for a formal firm is:

$$(1 + \tau_w) W n_{is} + R k_{is} + W c^F,$$

where c^F is the fixed cost, and τ_w is the labor tax. Taking input prices as given, standard cost minimization implies that the marginal cost of a firm is in the formal sector is:

$$mc_{is}^F = \frac{1}{z_{is}}\left(\frac{R}{\alpha}\right)^{\alpha}\left(\frac{W}{1-\alpha}\right)^{1-\alpha}(1+\tau_w)^{1-\alpha} = \frac{MC}{z_{is}}(1+\tau_w)^{1-\alpha}, \tag{3}$$

where $MC \equiv \left(\frac{R}{\alpha}\right)^{\alpha}\left(\frac{W}{1-\alpha}\right)^{1-\alpha}$ is the input price index, which is invariant across firms. Finally, we can write the profit maximization problem for formal firms as:

$$\begin{aligned} \pi_{is}^F = \max_{y_{is}, p_{is}} \quad & (1-\tau_y)p_{is}y_{is} - \frac{MC}{z_{is}}(1+\tau_w)^{1-\alpha}y_{is} - Wc^F \\ \text{s.t.} \quad & p_{is} = y_{is}^{-\frac{1}{\gamma}} y_s^{\frac{1}{\gamma}-\frac{1}{\theta}} Y^{\frac{1}{\theta}} P / M_s^{\frac{1}{\gamma}}, \end{aligned} \tag{4}$$

where τ_y is the sales tax. Within a market, there are strategic interactions, and the solution of the Cournot game implies that the optimal profit, price, and quantity of firm i depend on the optimal quantities of its $M_s - 1$ competitors.

The solution of the problem yields the usual pricing equation:

$$p_{is}^F = \underbrace{\frac{\varepsilon_{is}^F}{\varepsilon_{is}^F - 1}}_{\equiv \mu_{is}^F} \frac{MC}{z_{is}} \frac{(1+\tau_w)^{1-\alpha}}{1-\tau_y}, \tag{5}$$

where $\varepsilon_{is}^F > 1$ is the demand elasticity of a formal firm in market s and μ_{is}^F the associated markup. As in Atkeson and Burstein (2008), the demand elasticity is a weighted average of the within and across market elasticities:

$$\varepsilon_{is}^F = \left[\frac{1}{\gamma}(1-\omega_{is}^F) + \frac{1}{\theta}\omega_{is}^F\right]^{-1}, \tag{6}$$

where the weights, ω_{is}, are firm's revenue shares within its market: $\omega_{is} = p_{is}y_{is}/\sum_{i=1}^{M_s} p_{is}y_{is} = (p_{is}/p_s)^{1-\gamma}(1/M_s)$. As seen in Equation (6), when the firm is a monopolist, $\omega_{is} = 1$ and the only relevant elasticity is the across-market elasticity. If M_s is large and the market is very competitive, $\omega_{is} \to 0$, and the firm's elasticity of demand is given by the within-market elasticity.

Informal firms. Informal firms do not pay taxes or fixed costs, but they are subject to a size-dependent cost reflecting the increased likelihood of being observed by the tax authority and fined as they grow larger. The cost function, $C_I(z)$, is increasing in the firm's productivity, z, and is paid in labor units. The total cost for an informal firm is:

$$Wn_{is} + Rk_{is} + Wn_{is}C_I(z_{is}).$$

Assuming the following functional form for the size-dependent cost: $C_I(z) = \tau_1 z^{\tau_2} - 1$, where $\tau_1 \geq 1$ and $\tau_2 \in (0, 1)$, the marginal cost of a firm is in the informal sector is:

$$mc_{is}^I = \frac{1}{z_{is}} \left(\frac{R}{\alpha}\right)^{\alpha} \left(\frac{W}{1-\alpha}\right)^{1-\alpha} (\tau_1 z_{is}^{\tau_2})^{1-\alpha} = \frac{MC}{z_{is}^{(1-\tau_2(1-\alpha))}} \tau_1^{1-\alpha}, \tag{7}$$

and the profit maximization is:

$$\begin{aligned} \pi_{is}^I = \max_{y_{is}, p_{is}} \quad & p_{is} y_{is} - \frac{MC}{z_{is}^{(1-\tau_2(1-\alpha))}} \tau_1^{1-\alpha} y_{is} \\ \text{s.t.} \quad & p_{is} = y_{is}^{-\frac{1}{\gamma}} y_s^{\frac{1}{\gamma} - \frac{1}{\theta}} Y^{\frac{1}{\theta}} P M_s^{\frac{1}{\gamma}}. \end{aligned} \tag{8}$$

Again, the solution of the Cournot game implies the following pricing equation for informal firms:

$$p_{is}^I = \frac{\varepsilon_{is}^I}{\varepsilon_{is}^I - 1} \frac{MC}{z_{is}} \tau_1^{1-\alpha} z_{is}^{\tau_2(1-\alpha)} \tag{9}$$

where the markup, μ_{is}^I, and the demand elasticity, ε_{is}^I, are defined analogously to the formal counterparts. As seen from Equation (9), τ_1 acts as a uniform tax on labor, while τ_2 controls the size-dependent cost in the cost function.

Formalization decision. An individual firm decides to formalize if its profit in the formal sector is higher than its profits in the informal sector: $\pi_{is}^F \geq \pi_{is}^I$. As we fully discuss in Section 2.5, the formalization decision depends on idiosyncratic shocks (i.e., z_{is}), parameters (i.e., taxes and technology), as well as the aggregate quantity and price index. Importantly, the profit of a given firm will depend on the optimal quantities decided by its market competitors. This, in turn, depends on the competitors' own shocks and formalization decisions. The strategic interaction in production, combined with simultaneous formalization decisions, may lead to multiple equilibria, as different combinations of firms can decide to formalize, depending on the decisions of others. We discuss our equilibrium selection process in detail later.

2.3 Aggregation

Markets. Let ϕ_{is}^F be an indicator function that takes the value of one if the firm is formal and zero if it is informal. Define a sectoral markup μ_s such that it satisfies:

$$p_s = \mu_s \frac{MC}{Z_s} \frac{(1+\tau_w)^{(1-\alpha)}}{(1-\tau_y)}, \quad \text{where} \quad Z_s \equiv \left[M_s^{-1} \sum_{i=1}^{M_s} z_{is}^{\gamma-1} \right]^{\frac{1}{\gamma-1}}$$

is the market-efficient level of productivity given the measure of firms M_s. Using the market-level price index and the prices of the formal and informal firms, we can solve for the sectoral markups:

$$\mu_s = \frac{(1-\tau_y)}{(1+\tau_w)^{1-\alpha}} \left[\frac{(1+\tau_w)^{(1-\alpha)(1-\gamma)}}{(1-\tau_y)^{1-\gamma}} \sum_{i=1}^{M_s} \phi_{is}^F \left(\frac{z_{is}}{Z_s} \frac{1}{\mu_{is}^F} \right)^{\gamma-1} + \right.$$
$$\left. \dots \tau_1^{(1-\alpha)(1-\gamma)} \sum_{i=1}^{M_s} (1-\phi_{is}^F) \left(\frac{z_{is}^{1-\tau_2(1-\alpha)}}{Z_s} \frac{1}{\mu_{is}^I} \right)^{\gamma-1} \right]^{\frac{1}{1-\gamma}} \left(\frac{1}{M_s} \right)^{\frac{1}{1-\gamma}}. \tag{10}$$

The sectoral markup is shaped by two distinct distortions. First, dispersion in firm-level markups generated by market power reduces allocative efficiency and raises the sectoral price index. Second, informality entails a productivity penalty that directly affects firms' effective marginal costs. These distortions interact: because formal and informal firms charge different markups, informality mechanically increases markup dispersion within the sector. In addition, taxation further amplifies within-market dispersion in the presence of informality.

To compute the misallocation generated by markup dispersion and informality in a given market, let the market-level aggregate production function be:

$$Y_s = \Omega_s^{-1} Z_s (K_s)^\alpha (N_s^{\text{Oper.}})^{1-\alpha}, \tag{11}$$

where

$$N_s^{\text{Oper.}} = \sum_{i=1}^{M_s} n_{is} \qquad \text{and} \qquad K_s = \sum_{i=1}^{M_s} k_{is}$$

are the total labor in production (i.e., labor net of fixed and informality costs) and the total capital allocated in market s, respectively, and Ω_s^{-1} the wedge reflecting misallocation in the market (if > 1). After some calculations, we can show that:

$$\Omega_s = (\Omega_s^K)^\alpha (\Omega_s^N)^{1-\alpha} \tag{12}$$

where

$$\Omega_s^K \equiv \sum_{i=1}^{M_s} \left(\frac{z_{is}}{Z_s} \right)^{\gamma-1} \frac{\tau_{is}^{1-\alpha}}{M_s} \left(\frac{\phi_{is}^F \mu_{is}^F + (1-\phi_{is}^F)\mu_{is}^I (\tau_1 z_{is}^{\tau_2})^{(1-\alpha)} (1-\tau_y)/(1+\tau_w)^{(1-\alpha)}}{\mu_s} \right)^{-\gamma},$$

and

$$\Omega_s^N \equiv \sum_{i=1}^{M_s} \left(\frac{z_{is}}{Z_s} \right)^{\gamma-1} \frac{\tau_{is}^{-\alpha}}{M_s} \left(\frac{\phi_{is}^F \mu_{is}^F + (1-\phi_{is}^F)\mu_{is}^I (\tau_1 z_{is}^{\tau_2})^{(1-\alpha)} (1-\tau_y)/(1+\tau_w)^{(1-\alpha)}}{\mu_s} \right)^{-\gamma},$$

where $\tau_{is} = \phi_{is}^F(1+\tau_w) + (1-\phi_{is}^F)\tau_1 z_{is}^{\tau_2}$. As in Equation (10), the misallocation wedge increases due to markup dispersion and the informality penalty. Additionally, the term τ_{is} interacts with the informality decision and markups, further increasing misallocation. This term arises because the labor tax and informality cost distort the input decision (specifically, the capital-labor ratio).

Special cases. To build intuition, it is useful to analyze four special cases. First, consider a model with only labor as input (i.e., $\alpha = 0$).[3] In this case, the wedge collapses to

$$\Omega_s = \sum_{i=1}^{M_s} \frac{1}{M_s} \left(\frac{z_{is}}{Z_s}\right)^{\gamma-1} \left(\frac{\phi_{is}^F \mu_{is}^F + (1-\phi_{is}^F)\mu_{is}^I \tau_1 z_{is}^{\tau_2}(1-\tau_y)/(1+\tau_w)}{\mu_s}\right)^{-\gamma}. \tag{13}$$

In this scenario, since the input decision is not distorted, misallocation arises solely due to the dispersion in markups, the informality decision, and their interaction. Second, consider the case where all firms are formal ($\phi_{is}^F = 1$). In this situation, Equations (10) and (12) simplify to:

$$\mu_s^{\text{No inf.}} = \left[\frac{1}{M_s}\sum_{i=1}^{M_s}\left(\frac{z_{is}}{Z_s}\frac{1}{\mu_{is}^F}\right)^{\gamma-1}\right]^{\frac{1}{1-\gamma}}, \quad \text{and} \quad \Omega_s^{\text{No inf.}} = \sum_{i=1}^{M_s}\frac{1}{M_s}\left(\frac{z_{is}}{Z_s}\right)^{\gamma-1}\left(\frac{\mu_{is}^F}{\mu_s}\right)^{-\gamma},$$

which are exactly the expressions found in De Loecker et al. (2022). Third, if markups are constant (for instance, if the elasticity of substitution within and across markets are the same), $\tilde{\mu}_s = \mu_{is}^F = \mu_{is}^I$, the Equation (10) becomes:

$$\begin{aligned}\mu_s^{\text{Cnst. mkp}} =& \tilde{\mu}_s \frac{(1-\tau_y)}{(1+\tau_w)}\left[\frac{(1+\tau_w)^{1-\gamma}}{(1-\tau_y)^{1-\gamma}}\sum_{i=1}^{M_s}\phi_{is}^F\left(\frac{z_{is}}{Z_s}\right)^{\gamma-1} + \right. \\ & \left. \dots\ \tau_1^{1-\gamma}\sum_{i=1}^{M_s}(1-\phi_{is}^F)\left(\frac{z_{is}^{1-\tau_2}}{Z_s}\right)^{\gamma-1}\right]^{\frac{1}{1-\gamma}}\left(\frac{1}{M_s}\right)^{\frac{1}{1-\gamma}}\end{aligned} \tag{14}$$

Finally, consider the case where there is no dispersion in markups nor informality in sector s. In this case, $\mu_s = \tilde{\mu}_s$, there is no dispersion in productivity, and the sector-level productivity is at its efficient level, $\Omega_s = 1$. This holds true regardless of whether there are multiple inputs.

Final good. Finally, we can aggregate all markets to determine the aggregate markup and the aggregate wedge. Define the aggregate markup μ and the aggregate produc-

[3]See Online Appendix B for a model with only labor.

tion function such that they satisfy:

$$P = \mu \frac{MC}{Z} \frac{(1+\tau_w)^{1-\alpha}}{(1-\tau_y)}, \quad \text{and} \quad Y = \Omega^{-1} Z (K)^{\alpha} (N^{\text{Oper.}})^{1-\alpha}, \tag{15}$$

where

$$Z \equiv \left[\int_0^1 Z_s^{\theta-1} ds \right]^{\frac{1}{\theta-1}}, \quad K \equiv \int_0^1 K_s ds, \quad \text{and} \quad N^{\text{Oper.}} \equiv \int_0^1 N_s^{\text{Oper.}} ds.$$

Following the same steps as before, the aggregate markup is:

$$\mu = \left[\int_0^1 \left(\frac{Z_s}{Z} \frac{1}{\mu_s} \right)^{\theta-1} ds \right]^{\frac{1}{1-\theta}}, \tag{16}$$

and the aggregate wedge is:

$$\Omega = \left[\int_0^1 \Omega_s^K \left(\frac{Z_s}{Z} \right)^{\theta-1} \left(\frac{\mu_s}{\mu} \right)^{-\theta} ds \right]^{\alpha} \left[\int_0^1 \Omega_s^N \left(\frac{Z_s}{Z} \right)^{\theta-1} \left(\frac{\mu_s}{\mu} \right)^{-\theta} ds \right]^{1-\alpha}. \tag{17}$$

Again, the aggregate wedge depends on the dispersion of sectoral markups (μ_s) and market-level wedges (Ω_s^K and Ω_s^N). This dispersion is influenced by the extent to which different markets exhibit varying degrees of informality and market power. The greater the dispersion, the larger the aggregate wedge ($\Omega > 1$) will be, resulting in lower aggregate output. When there is no dispersion in market markups or wedges, prices align with marginal costs, and the aggregate wedge equals one, indicating no misallocation.

2.4 Market Clearing and Equilibrium Algorithm

We solve for the equilibrium in the steady state. As seen from Equations (5) and (9), the prices and aggregates only enter the pricing equation through the input price index, MC. Since the market share is written as $\omega_{is} = (p_{is}/p_s)^{1-\gamma}$, the input price index cancels out when plugging in the pricing equations, rendering the market share independent of any aggregates. Therefore, the market equilibrium is *block recursive*. Given a vector of firms' productivities in market s, $(z_{1s}, z_{2s}, ..., z_{M_s s})$, and a vector of formalization decisions, $(\phi_{1s}, \phi_{2s}, ..., \phi_{M_s s})$, we can solve for the market shares, markups, and demand elasticities of each firm in the economy. Furthermore, by aggregating firm-level markups, we can solve for market-level (μ_s) and economy-wide markups (μ), as well as wedges (Ω_s^K, Ω_s^N, Ω_s, and Ω).

Once we have the economy-wide markup, we can use the market-clearing conditions to solve for the general equilibrium. As usual, the Euler equation in the steady state determines the interest rate, R. Then, from the aggregate pricing equation (i.e., Equation (15)), there is a one-to-one relationship between the wage rate, W, and the aggregate price index, P. With a normalization of the wage rate or the aggregate price index, we can use the labor market clearing condition to find the aggregate output, Y. We can express the labor demand equation as a sum of three components:

$$N^d = \underbrace{\int_0^1 \sum_{i=1}^{M_s} n_{is} ds}_{\equiv N^{\text{Oper.}}} + \underbrace{\int_0^1 \sum_{i=1}^{M_s} \phi_{is}^F c^F ds}_{\equiv N^{\text{Fixed Cost}}} + \underbrace{\int_0^1 \sum_{i=1}^{M_s} (1-\phi_{is}^F) n_{is} C_I(z_{is}) ds}_{\equiv N^{\text{Inf. Cost}}} \tag{18}$$

where $N^{\text{Oper.}}$ is the total labor used in production, $N^{\text{Fixed Cost}}$ is the labor used to pay for the fixed cost, and $N^{\text{Inf. Cost}}$ is the labor used to cover the informality cost. Note that n_{is} is a decreasing function of W and an increasing function of P and Y. In equilibrium, the labor demand equals the household labor supply:

$$N^s = \hat{\nu} \left(\frac{W}{P}\right)^{\nu}. \tag{19}$$

Thus, the set of equations used to determine the equilibrium Y, W, P (up to a normalization) are:

$$P = \mu \frac{MC}{Z} \frac{(1+\tau_w)^{1-\alpha}}{(1-\tau_y)} \qquad \text{and} \qquad N^s\left(\frac{W}{P}\right) = N^d(Y, W, P). \tag{20}$$

The Online Appendix D provides the procedure used to compute the general equilibrium solution of the model.

The general equilibrium solution described above assumes the firms' formalization decisions are given. We must then check if the formalization decisions of all firms are consistent with the general equilibrium. As mentioned before, strategic interactions combined with simultaneous formalization decisions may lead to multiple equilibria. We select the equilibrium using a simple algorithm as follows: First, we solve for the market equilibrium and the market clearing conditions starting with all firms in the formal sector. Second, we rank the firms by their profits in the formal sector.[4] Then, we check if it is profitable for the formal firm with the lowest profit in each market to move to the informal sector. If it is, we move this firm to the informal sector and compute the market equilibrium and the aggregate market clearing conditions again.

[4]Since both productivities and fixed costs are stochastic, the most profitable firms are not necessarily the most productive.

This procedure is repeated until no firm moves to the informal sector.

2.5 The Role of Informality in Shaping Market Power

In our model, as in Edmond et al. (2023), markup dispersion affects equilibrium through multiple channels. The first channel is the increase in the aggregate markup, μ, which directly reduces real wages and labor supply, as seen in Equation (20). The second channel is the misallocation channel, which appears in the aggregate wedge, Ω. Additionally, variable markups influence the formalization decision through changes in profits. However, unlike previous models in the literature, the inclusion of an additional distortion — namely, informality — interacts with the dispersion of markups.

Given the potential interaction between informality and market power, is such an interaction positive or negative? That is, does informality reduce the distortions generated by market power, or does it amplify them? The answer depends on many factors, including the magnitude of sales taxes, corporate taxes, fixed costs, and the productivity of informal firms. In this section, we illustrate how informality shapes market power in our model.

According to equation (6), demand elasticity, and thus the inverse markup, is directly related to a firm's market share. Consequently, the revenue share serves as a sufficient statistic for estimating a firm's market power in a given market. The relevant question then becomes whether a firm's decision to become informal increases or decreases its revenue share. If informality increases the revenue share of an informal firm, it could counteract the market power by reducing the revenue share of other, potentially larger firms. To explore this, it is useful to compare the revenue functions of formal and informal firms as functions of their productivity and market-level aggregates. The revenue function of a formal firm with productivity z_{is}, demand elasticity ε^F_{is} in market s is

$$R^F(z_{is}, \varepsilon^F_{is}, y_s, p_s) = \left(\frac{\varepsilon^F_{is} - 1}{\varepsilon^F_{is}} \frac{z_{is}}{MC} \frac{(1-\tau_y)}{(1+\tau_w)^{1-\alpha}} p_s \right)^{\gamma-1} \frac{p_s y_s}{M_s}, \tag{21}$$

while the revenue function of an informal firm is

$$R^I(z_{is}, \varepsilon^I_{is}, y_s, p_s) = \left(\frac{\varepsilon^I_{is} - 1}{\varepsilon^I_{is}} \frac{z_{is}^{1-\tau_2(1-\alpha)}}{MC} \frac{1}{\tau_1^{1-\alpha}} p_s \right)^{\gamma-1} \frac{p_s y_s}{M_s}. \tag{22}$$

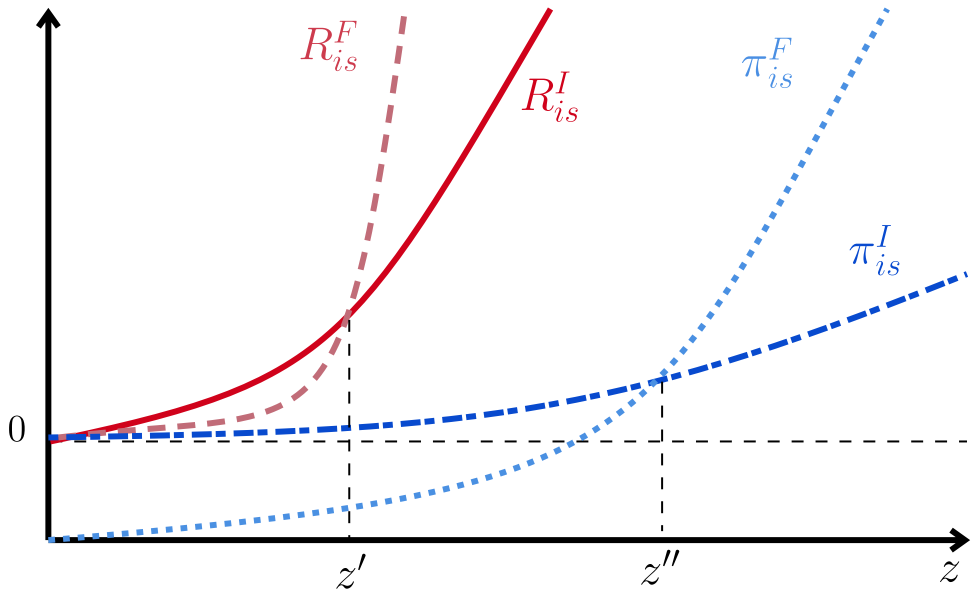
According to equations (21) and (22), providing that $(1+\tau_w)^{1-\alpha}/(1-\tau_y) > \tau_1^{1-\alpha}$, the revenue function of an informal firm will be larger at low productivity levels. However, as productivity increases, the revenue function of formal firms grows faster, eventually surpassing that of informal firms. This intersection occurs at z' in Figure 1. Thus, if a firm with productivity greater than z' opts for informality, its revenue will decrease, thereby increasing the market power of other firms in the market.

The decision to formalize depends on the profits firms can achieve in each sector. The profit functions for formal and informal firms are:

$$\pi_{is}^F = (1-\tau_y)\frac{R^F(z_{is}, \varepsilon_{is}^F, y_s, p_s)}{\varepsilon_{is}^F} - Wc^F \qquad \text{and} \qquad \pi_{is}^I = \frac{R^I(z_{is}, \varepsilon_{is}^I, y_s, p_s)}{\varepsilon_{is}^I}.$$

These profit functions are themselves functions of the revenue functions. However, the presence of sales taxes and fixed operational costs allows for scenarios where firms have higher profits in the informal sector despite having lower revenue. For such firms, remaining informal is the optimal choice, thereby increasing the market power of larger firms.

Figure 1: Revenue and Profit Functions of Formal and Informal Firms



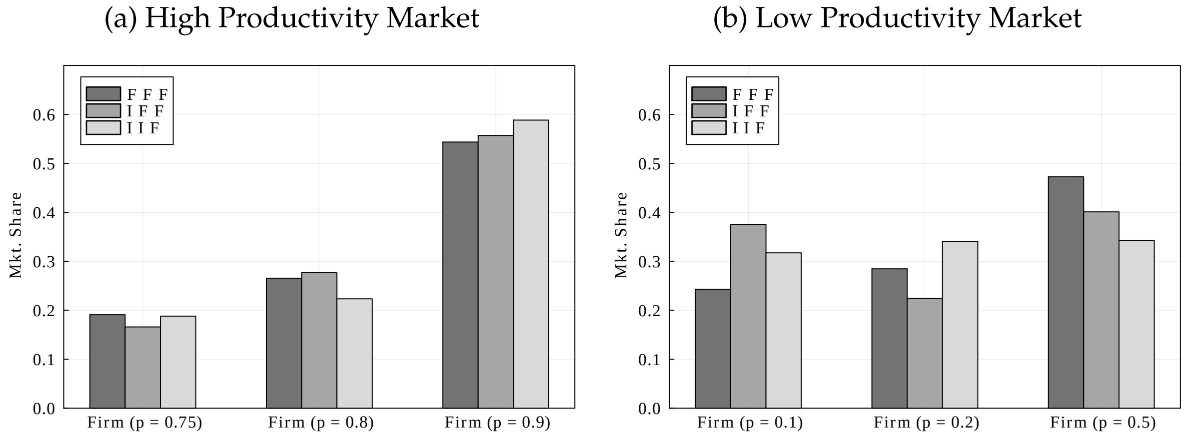
Notes: The figure plots the revenue and profit functions for a fixed demand elasticity and fixed cost in a given market s. R_{is}^F and R_{is}^I denote the revenue functions of formal and informal firms, while π_{is}^F and π_{is}^I denote the profit functions of formal and informal firms.

Figure 1 illustrates the profit functions for a fixed demand elasticity, i.e. $\varepsilon_{is}^F = \varepsilon_{is}^I$.[5] The decision to formalize follows a cutoff rule in the productivity space. Firms with productivity to the right of z'' will achieve higher profits in the formal sector and will therefore formalize, while those to the left will remain informal. Due to sales taxes and fixed costs, the intersection point, z'', between the profit functions is to the right of the intersection point z' between the revenue functions. Firms within the interval $z \in (z', z'')$ will opt to be informal, increasing the market power of larger firms. The

[5]Since the market share of a firm changes with its formalization decision, the demand elasticity does not need to be same.

larger the sales tax and fixed costs, the greater the range of firms with higher informal profits but lower revenues. Finally, firms with productivity below z' are larger in the informal sector. For these firms, informality reduces the market power of larger firms.

Figure 2: Three-firm Markets with Different Levels of Informality



Notes: Firm p means firm with productivity in the percentile of the productivity distribution.

The demand elasticity and the market shares are endogenous to the formalization decision of the firms, but still the main point remains. The question of whether informality decreases or increases the market power depends on how many firms are in each region of the productivity space. Figure 2 illustrates this point in two markets with three firms. In the high-productivity market (left panel), there are firms in the 75th, 80th, and 90th percentiles of the productivity distribution. We consider three cases (each bar represents a case): in the first case all three firms are formal, in the second case the least productive firm is informal, and in the third case only the most productive firm is formal. The market share of the largest firm increases gradually, as more firms become informal. This is the case as these firms have higher revenue in the formal than in the informal sector. In the right panel, we perform the same experiment in a low-productivity market. In this market, the three firms are in the 10th, 20th, and 50th percentiles of the productivity distribution. As we move the low-productivity firms to informality, the market share of the leader *decreases*. In this market, the least productive firms increase revenue when moving to informality. There are many other possible types of market structures (e.g., markets with the same distribution of productivity but different numbers of firms, markets where very high and low-productivity firms co-exist, etc). Ultimately, whether informality increases or decreases market power is a quantitative question.

3 Calibration

We calibrate the model by matching moments from the Brazilian data. In particular, we use a combination of multiple data sets to derive moments for the formal and informal sectors. Our markup and sales data come from PIA (*Pesquisa Industrial Anual*), an administrative firm-level data set for the manufacturing sector that covers the universe of large firms and a representative sample of small firms. We also use the matched employer-employee data set RAIS (*Relação Anual de Informações Sociais*), which includes the universe of all employment contracts for formal firms. Using RAIS, we calculate firm size and employment share among formal firms. The sales and size data from informal firms come from the ECINF (*Economia Informal Urbana*). Finally, we supplement aggregate informality employment from the nationally representative household survey, PNAD (*Pesquisa Nacional por Amostra de Domicílios*).

Before we proceed with the calibration, we discuss the functional forms. Then, we describe the moments used in the estimation and the intuition behind the identification of the parameters.

3.1 Functional Forms

In the quantitative exercise, we assume that the individual firm's productivity is the product of an idiosyncratic component, $\tilde{z}_{is}$, and a market-level component, a_s: $z_{is} = a_s\tilde{z}_{is}$. The market-specific component $a_s \geq 1$ is drawn from a log-normal distribution truncated below at one, $a_s \sim \text{Lognormal}(-\sigma_a^2/2, \sigma_a^2)$, while the idiosyncratic component $\tilde{z}_{is} \geq 1$ is drawn from a log-normal distribution truncated below at one, $\tilde{z}_{is} \sim \text{Lognormal}(-\sigma_z^2/2, \sigma_z^2)$. Both components are drawn i.i.d. across producers and sectors.

Following Bao et al. (2022), we assume that the number of firms in a market, $M_s \in \mathbb{N}$, is stochastic and drawn from a normal distribution, $N(\mu_m, \sigma_m^2)$. Specifically, we first draw from the continuous distribution truncated to the interval $(0, \overline{M}]$, add one, and then round the result up to the nearest integer.[6] Finally, we allow the market-specific productivity and the number of firms in a market to be correlated. We model the dependence between these two distributions ($N(\mu_m, \sigma_m^2)$ and $\text{Lognormal}(-\sigma_a^2/2, \sigma_a^2)$) using a Gaussian copula with correlation parameter ρ.

[6] We set $\overline{M} = 200$.

3.2 Parametrization

Exogenously calibrated parameters. The exogenously calibrated parameters are reported in Table 1. We set labor and sales tax rates equal to the statutory values in 2003 (see Ulyssea (2018) for a discussion). The capital share and the depreciation rate are set to standard values, $\alpha = 1/3$ and $\delta = 0.14$. The discount factor β is chosen to match a real interest rate of 6%, implying a rental rate of capital of $R = 0.2$. The aggregate price index P is normalized to one. Following De Loecker et al. (2022), we set the labor supply elasticity equal to 0.25, a conservative value according to Chetty et al. (2011). A higher labor supply elasticity amplifies the negative effects of markups according to Equation (19). Finally, we must choose the within-market and across-market elasticities of substitution. To build intuition for their role in shaping market power in the model, note that they directly determine the upper and lower bounds of firm markups. When the number of firms in a market is large, an individual firm's market share is negligible and, by Equation (6), its markup will be equal to $\mu_{is} = \gamma/(\gamma - 1)$. In contrast, when a firm is the sole producer in a market, its markup is $\mu_{is} = \theta/(\theta - 1)$. We set $\theta = 1.50$ and $\gamma = 10.0$, implying a lower bound on markups of 1.11 and an upper bound of 3.0.[7] These values fall within the range commonly used in models with a nested CES production structure similar to ours.[8]

Table 1: Exogenously Calibrated Parameters

Parameter		Value	Source
Within-market elasticity	γ	10.00	See text
Between-market elasticity	θ	1.50	See text
Labor supply elasticity	ν	0.25	Chetty et al. (2011)
Sales tax	τ_y	0.29	Ulyssea (2018)
Labor tax	τ_w	0.375	Ulyssea (2018)
Capital share	α	1/3	Standard Values
Rental rate of capital	R	0.2	Standard Values
Agg. price index	P	1.00	Normalization

Internally estimated parameters. We choose the remaining parameters, $\Theta = \{\tau_1, \tau_2, c^F, \mu_m, \sigma_m, \sigma_z, \sigma_a, \rho\}$, to minimize the distance between model-simulated moments and their empirical counterparts. Table 2 reports the resulting parameter values.[9] In particular, we target a set of moments capturing both the levels and disper-

[7] In the data, many firms have markups below unity. We interpret this as firms incurring losses in anticipation of higher future profits. Since our model is static, it cannot capture these dynamics.

[8] For example, estimates of the within-market elasticity range from 5 (Burstein et al., 2023) to as high as 10.5 (Edmond et al., 2015), while estimates of the across-market elasticity typically lie between 1.2 and 1.8.

[9] The current calibration is preliminary and might change in future iterations of the paper.

sions of markups, sales, and employment in the formal and informal sectors. To compute model moments, we simulate 10,000 markets by drawing from the distributions of productivity and the number of firms. As discussed above, the number of firms in each market, M_s, is stochastic and follows a normal distribution.

Table 2: Internally Estimated Parameters

Parameter		Value
Level cost of informality	τ_1	2.151
Convex cost of informality	τ_2	0.221
Formal fixed cost	c^F	0.000065
Mean number of firms in a market	μ_m	54.0
Dispersion of number of firms in a market	σ_m	51.0
Dispersion of firm-level productivity distribution	σ_z	0.32
Dispersion of market-level productivity distribution	σ_a	0.35
Dependence of market productivity and number of firms	ρ	-0.6585

3.3 Moments and Identification

Table 3 reports the model's fit to the targeted moments. We group the moments into (i) markup moments and (ii) firm and employment distribution moments. The markup moments include the sales-weighted average markup, the between-sector standard deviation of markups, and the within-sector standard deviation of markups. These moments are computed from PIA data, are sales-weighted, and include only formal firms. We impose the same restrictions when computing the corresponding model-simulated moments.

Table 3: Data and Model Moments

Moments	Data	Model	Source
Average markup (formal firms and sales weighted)	1.18	1.22	PIA
Between-mkt markup std. (formal firms and sales weighted)	0.077	0.077	PIA
Within-mkt markup std. (formal firms and sales weighted)	0.091	0.089	PIA
Share of informal firms among all firms	0.700	0.746	RAIS + ECINF
Share of informal workers	0.350	0.351	PNAD
Ratio of avg. revenue among small firms	6.40	6.21	ECINF
Share of small firms among all inf. firms	0.998	0.978	ECINF
Partial corr. between markups and informality	-0.400	-0.397	PNAD + PIA

The firm and employment distribution moments capture the share of informal firms, the share of small firms among informal firms, the share of informal workers in the labor force, and the ratio of average revenue between small formal firms and informal firms. These moments are constructed using a combination of RAIS, ECINF, and

PNAD data.[10] Finally, we target the sector-level partial correlation between markups and informality. Specifically, in both the data and the model, we estimate a regression of the sectoral share of informal workers on sectoral markups and include the slope coefficient as a targeted moment.

While all parameters are jointly chosen to match the targeted moments, different moments are more informative about specific parameters. This section summarizes the main identification channels. A first set of moments disciplines market structure and productivity dispersion. In the model, market power is decreasing in the number of firms: markets with more firms exhibit systematically lower average markups. As a result, the sales-weighted average markup is tightly linked to the mean number of firms per market, μ_m. Moreover, because profits are lower in markets with many competitors, a larger fraction of firms chooses informality. Variation in the number of firms across markets therefore, affects both the level and the dispersion of markups. In particular, within- and between-sector markup dispersion jointly identify dispersion in firm-level productivity, σ_z, and market-level productivity, σ_a.

A second set of moments identifies the costs of formality and informality. The formal fixed cost, c^F, and the level and convex components of informality costs, τ_1 and τ_2, are determined by moments of the firm size and employment distributions. These parameters affect informality and firm size along distinct margins, allowing the share of informal firms, the share of informal workers, and the ratio of average revenues between small formal and informal firms to jointly identify them. An increase in the formal fixed cost raises informality by pushing marginal firms into the informal sector. Because these marginal firms are larger than the average informal firm, this reallocation increases the average size of informal firms and, mechanically, the average size of formal firms as well. By contrast, a reduction in the level cost of informality shifts marginal firms into informality and also expands the scale of inframarginal informal firms. A reduction in the convex cost of informality has a similar effect, but because the cost is productivity-dependent, it generates heterogeneous size responses across the productivity distribution. Finally, these parameters also affect formal-sector markups indirectly by changing both the revenue share of formal firms and the composition of firms that remain formal.

The partial correlation between formal-sector markups and informality identifies the dependence between market productivity and the number of firms, ρ. More productive markets tend to exhibit lower informality because high-productivity firms can more easily cover the fixed cost of the formal sector. At the same time, lower in-

[10] In the data, a small firm is defined as having five or fewer employees. In the formal sector, approximately 70% of firms are small. In the model, we therefore define small firms as those below the 70th percentile of the formal-sector size distribution.

formality implies that smaller firms, relative to the sectoral distribution, and with lower markups, operate in the formal sector, putting downward pressure on average markups. Finally, markups decline mechanically with the number of firms in the market. Absent a sufficiently negative dependence between market productivity and the number of firms, these forces offset each other, and the model fails to reproduce the negative partial correlation between markups and informality observed in the data.

Non-targeted moments. The model also performs well along non-targeted dimensions. Table 4 reports the fit. In particular, the model closely matches the aggregate labor share and generates a modestly higher degree of concentration in the formal sector, as measured by the national Herfindahl–Hirschman Index (HHI), than observed in the data. In addition, the model implies lower markups in the informal sector, consistent with the notion that informality is associated with more competitive pricing and lower market power.

Table 4: Non-targeted Moments

Moments	Data	Model	Source
Labor share	0.534	0.556	Penn World Table
National HHI (formal, sales weighted)	0.110	0.152	PIA
Avg. markup (inf. firms and sales weighted)	-	1.170	-

4 Quantitative Results

This section presents the quantitative results. We first assess the aggregate role of market power in Brazil and then use the model to conduct a series of counterfactual exercises that quantify how informality, market power, and taxation interact. Before turning to the results, we define some key variables used in the analysis.

Throughout the counterfactual exercises, we measure national markups in the same way as in the data, using the sales-weighted average of firm-level markups.[11]

$$\overline{\mu} = \int \sum_{i=1}^{M_s} \left(\frac{p_{is} y_{is}}{PY} \right) \mu_{is} ds.$$

As discussed in Section 2.5, the aggregate cost of market power can be decomposed into two components: the level effect, captured by μ, and the misallocation wedge,

[11]Note this measure differs from the harmonic-mean aggregate markup, μ, that arises from the model.

captured by Ω. The former reflects the average markup *level*, while the latter summarizes the efficiency losses arising from markup *dispersion*.

From the aggregate pricing equation (15), a higher aggregate markup μ reduces the real wage. Through the labor supply condition (Equation 20), this implies a lower equilibrium labor supply N. Let $\hat{N}$ denote the counterfactual labor supply when $\mu = 1$, holding all other objects fixed. Define operational labor in this counterfactual as

$$\hat{N}^{Oper.} = \hat{N} - N^{\text{Fixed Cost}} - N^{\text{Inf. Cost}}.$$

Holding capital and informality fixed, counterfactual output when $\mu = 1$ is

$$\hat{Y} = Z\Omega^{-1}(K)^{\alpha}(\hat{N}^{Oper.})^{1-\alpha}.$$

Similarly, counterfactual output when the misallocation wedge is shut down ($\Omega = 1$) is

$$\tilde{Y} = \Omega Y.$$

Finally, when both distortions are removed ($\mu = 1$ and $\Omega = 1$), output is

$$Y^{*} = \Omega \hat{Y}.$$

Note that these decompositions are accounting exercises rather than full general-equilibrium counterfactuals: capital accumulation and informality decisions remain fixed. They therefore reflect the partial-equilibrium contribution of markup levels and dispersion to aggregate output.

4.1 Inspecting the Baseline Results

We are now ready to discuss the main results. Table 5 reports the aggregate statistics of the baseline model. Informal firms account for 74.5% of all firms, yet they employ only 34.8% of workers and generate 20.9% of aggregate output. Informality is therefore pervasive along the extensive margin but quantitatively modest in terms of production. Consistent with their small scale, informal firms charge lower markups: 1.17 on average, compared to 1.22 in the formal sector. This difference reflects not only their smaller size but also the fact that they tend to operate in more competitive markets with a larger number of firms.

Aggregating across firms yields an economy-wide markup of $\mu = 1.238$. In addition to this level effect, markup dispersion generates a misallocation wedge of $\Omega = 1.031$, implying an efficiency loss of about 3.1%. The output decomposition highlights the

Table 5: Aggregate Results

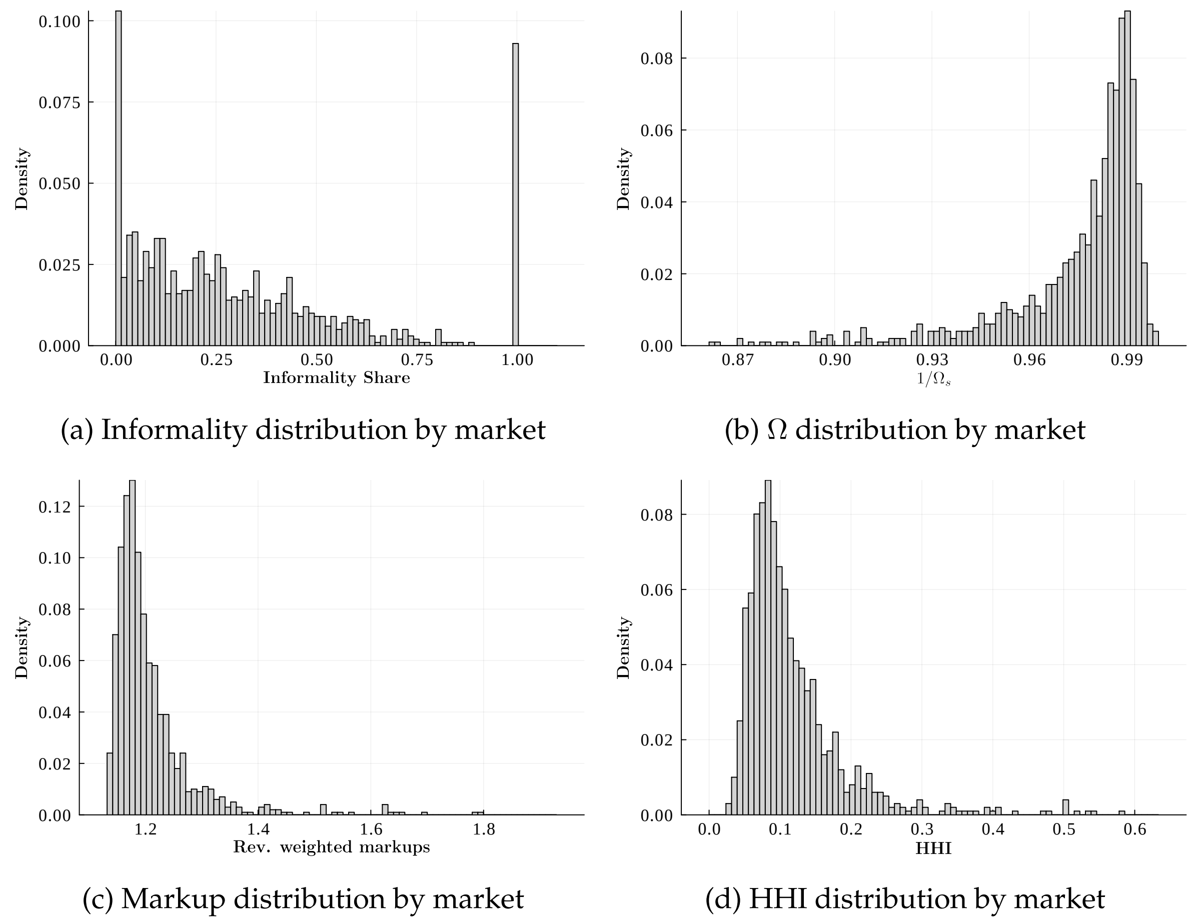
	Baseline	Low (half) Informality	No Informality
Output - Relative to Baseline	1.000	1.105	1.213
Real Wage - Relative to Baseline	1.000	1.004	1.010
Share Informal Firms	0.745	0.621	0.000
Share Informal Workers	0.348	0.178	0.000
Informal Sales Share	0.209	0.098	0.000
Tax Revenue / GDP	0.314	0.358	0.398
Avg. Markup Formal	1.221	1.215	1.211
Avg. Markup Informal	1.170	1.179	-
μ	1.238	1.235	1.230
Ω	1.031	1.028	1.025
$\hat{Y}/Y$	1.045	1.040	1.035
$\tilde{Y}/Y$	1.031	1.028	1.025
Y^*/Y	1.078	1.069	1.061

Notes: Markups are revenue-weighted averages. $\hat{Y}$ denotes output when $\mu = 1$; $\tilde{Y}$ denotes output when $\Omega = 1$; and Y^* denotes output when both $\mu = 1$ and $\Omega = 1$.

relative importance of these two channels. Eliminating the level distortion ($\mu = 1$) raises output by 4.5%, while removing the dispersion wedge ($\Omega = 1$) increases output by 3.1%. Eliminating both distortions simultaneously increases output by 7.8%. Recall that these calculations hold the capital stock and informality decisions fixed.

Market-level heterogeneity. The aggregate statistics mask substantial heterogeneity across markets. Figure 3 reports the cross-market distributions of informality, average markups, concentration (HHI), and the misallocation wedge (Ω_s). Panel (a) shows wide dispersion in informality across markets. Roughly 10% of markets exhibit no informality, while slightly fewer than 10% are fully informal. The remaining markets span the entire interior of the distribution, indicating that informality is not uniformly distributed but highly market-specific. Concentration also varies considerably. As shown in Panel (d), most markets have an HHI between 0.06 and 0.15; yet, a non-negligible mass exhibits an HHI above 0.5, indicating highly concentrated environments. This variation in concentration is mirrored in the distribution of revenue-weighted markups. While the bulk of markets display average markups around 1.2, a subset reaches values as high as 1.8. The interaction between informality and market power generates substantial dispersion in the market-level wedge Ω_s. For most markets, misallocation reduces output by 1–3%. However, a meaningful fraction of markets experience output losses exceeding 10%. Hence, aggregate misallocation is driven not only by average distortions, but also by a tail of severely distorted markets.

Figure 3: Informality, Markups, and Concentration Across Markets



(a) Informality distribution by market

(b) Ω distribution by market

(c) Markup distribution by market

(d) HHI distribution by market

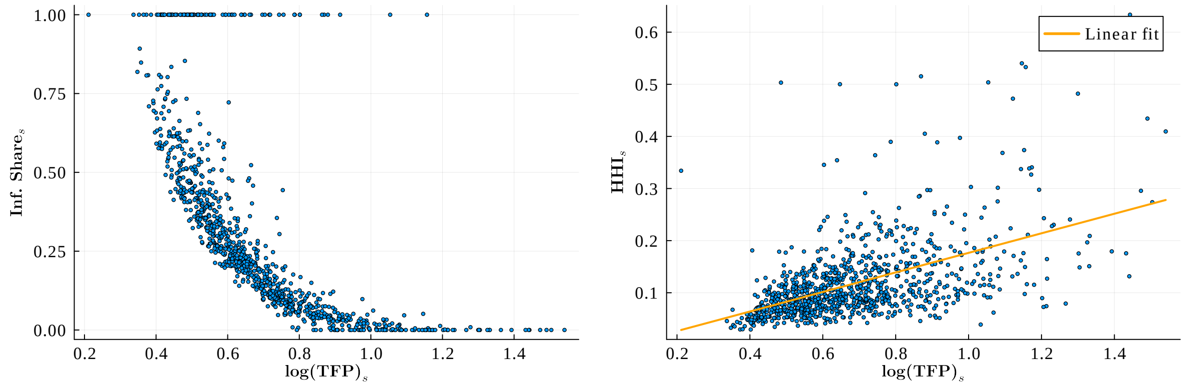
Notes:

To study how distortions relate to productivity, Figure 4 plots market-level informality and concentration against log TFP, where market TFP equals $Z_s\Omega_s$. The left panel shows a strong, non-linear negative relationship between informality and TFP. Markets with high informality systematically exhibit lower productivity. This reflects two reinforcing mechanisms. First, informality directly reduces TFP through the misallocation wedge. Second, low-productivity firms disproportionately select into informality, amplifying the aggregate productivity loss. Informality, therefore, both reflects and exacerbates low productivity. The right panel shows a positive correlation between HHI and TFP. This relationship is more subtle. On the one hand, lower concentration reduces markup dispersion and improves efficiency. On the other hand, high-productivity markets tend to have fewer firms and exhibit lower informality, mechanically increasing concentration.[12] To isolate this composition effect, Figure A.1 in the Appendix plots the correlation between TFP and HHI computed within the formal

[12]Recall that the negative partial correlation between markups and informality was a target moment in our calibration.

sector only. Once informal firms are excluded, the positive correlation largely disappears, suggesting that the baseline relationship is primarily driven by selection and composition rather than a direct effect of concentration.

Figure 4: Productivity and Market Structure Across Markets



Notes:

Decomposition of the aggregate misallocation wedge. We further decompose the misallocation wedge arising from markup dispersion (Ω) using Equations (17) and (12). Specifically, we evaluate counterfactual wedges by sequentially shutting down: (i) informality; (ii) dispersion of markups within markets; and (iii) dispersion of markups between markets. Table 6 reports the resulting values of Ω and μ in each case.

Table 6: Decomposition of Ω

	Ω	μ
Baseline	1.032	1.239
No Informality	1.025	1.231
No Dispersion Within	1.011	1.247
No Dispersion Within + No Inf.	1.005	1.239
No Dispersion Within and Between	1.006	1.247
No Dispersion Within and Between + No Inf.	1.000	1.239

Notes:

In the "No Informality" scenario, we keep firm-level markups at their baseline values but set all firms to operate formally (i.e., $\phi_{is}^F = 1$ for all i). This reduces the wedge from 1.032 to 1.025, a decline of 0.7 percentage points. Next, we eliminate dispersion of markups within markets by replacing firm-level markups with their corresponding market-level averages. This reduces Ω to 1.011, a decline of 2.1 percentage points relative to the baseline. When we additionally eliminate informality, the wedge falls further to 1.005. Finally, shutting down all markup dispersion, by setting firm-level markups equal to the aggregate markup ($\mu_{is} = \mu$), removes both within- and between-market dispersion. In this case, the wedge declines to 1.006, and to 1.0 once informality is also eliminated. Overall, roughly two-thirds of the baseline misallocation arises from dispersion of markups within markets, with a smaller contribution from between-market dispersion and informality.

4.2 The Role of Informality in Shaping Market Power

Section 2.5 showed that the interaction between informality and market power is theoretically ambiguous. Here we quantify that interaction. In the model, a firm's markup depends on its residual demand, which in turn depends on the distribution of market shares within the market. Formalizing informal firms therefore affects markups of formal firms through the changes it induces in market shares. Figure 1 shows that there exists a productivity threshold z' at which the revenue functions of formal and informal firms intersect. For firms with productivity below z', formalization reduces their market share; for firms above z', formalization increases it. Holding everything else fixed, the overall effect of eliminating informality on market power depends on how firms are distributed across these regions of the productivity space.

Table 7: Fraction of firms and market share by productivity

Productivity Space	$[z_{min}, z')$	$[z', z'')$	$[z'', z_{max})$
Share of Firms (avg. per mkt) (1)	0.262	0.400	0.338
Mkt. Share (avg. per mkt) (2)	0.033	0.193	0.773
Mkt. Share No Inf (avg. per mkt) (3)	0.027	0.216	0.757
Change (3) - (2)	-0.006	0.022	-0.016

Table 7 reports the average fraction of firms and their market shares across productivity intervals. About 26% of firms lie below z', 40% between z' and z'', and 34% above z''. Although there are many low-productivity firms, they account for only 3.3% of total production. In contrast, the highest-productivity (formal) firms account for 77.3%. We then conduct a partial-equilibrium counterfactual in which informal firms are formalized and evaluated under the formal-sector revenue function, keeping markups and all other equilibrium objects fixed. Market shares adjust mechanically. The share of low-productivity firms falls from 3.3% to 2.7%, while the share of intermediate-productivity firms rises from 19.3% to 21.6%. The overall change implies a modest 1.6 percentage point decline in the market share of the most productive firms.

Two forces explain why the aggregate effect is small. First, even though there is a large number of informal firms, they account for little output, so changes in their shares have a limited impact on the residual demand of their competitors. Second, the shifts across productivity regions partially offset each other: some firms gain share upon formalization while others lose it. Informality thus has non-trivial distributive consequences across firms, but only a modest effect on aggregate market power in this partial-equilibrium exercise.

In the previous exercise, we held all endogenous variables fixed. A natural next step is to allow firms to reoptimize, capital to adjust, and markets to clear in general equilibrium. Table 5 reports the results for two counterfactuals: a "No Informality" scenario, in which firms are not allowed to operate informally, and a "Low Informality" scenario, in which we increase τ_1 to target half of the baseline share of informal workers (17.8%).

Reducing informality has large aggregate effects. Output increases by 10.5% in the low-informality economy and by 21.3% under full formalization. Real wages rise modestly - by 0.4% and 1%, respectively - while tax revenue increases substantially, reaching 35% and 39% of GDP. Despite these improvements, market power continues to generate non-trivial distortions. Setting $\mu = 1$ still raises output by 4.0% and 3.5% in the two counterfactuals, while eliminating the dispersion wedge ($\Omega = 1$) accounts for an additional 2.8% and 2.5%. Notably, the aggregate wedge Ω in the no-informality economy coincides with the partial-equilibrium decomposition in Table 6, reinforcing the earlier finding that eliminating informality has only a modest direct effect on markup dispersion.[13]

Table 8: Relative Output by Group of Firm

	Baseline	Low (half) Informality	No Informality
Relative Y always formal	1.000	1.096	1.191
Share of always formal	—	0.255	0.255
Relative Y always informal	1.000	0.965	—
Share of always informal	—	0.621	0.000
Relative Y inf to formal	1.000	1.331	1.295
Share of inf to formal	—	0.123	0.745
Relative Y formal to informal	1.000	—	—
Share of formal to informal	—	0.000	0.000

Notes:

What drives the large increase in output? To answer this, Table 8 decomposes the

[13] Appendix Table A.1 reports the analogous decomposition for the low-informality economy.

results by firm group in the low-informality scenario.[14] Three groups are relevant.

First, firms that are formal both in the baseline and in the counterfactual account for about 26% of firms and most of aggregate production. Their output rises by 9.6%. Two forces are at work. On the one hand, aggregate output expands, increasing market-level demand. On the other hand, their markups decline as competitors, particularly those transitioning from informality, expand more aggressively. To preserve market share, these firms increase production despite paying higher wages. Second, firms that remain informal (62% of firms in the low-informality economy) reduce output by 3.5%. For these firms, the increase in the real wage dominates the market-level expansion in demand. Given their small scale and limited ability to adjust markups, they contract. Finally, 12.3% of firms transition from informal to formal status. This group increases output by 33%. Formalization relaxes distortions in their revenue function and allows them to operate at a larger scale. These firms, roughly middle-sized in the productivity distribution, account for a significant portion of the aggregate output gain.

Overall, the increase in output reflects both the direct expansion of firms that formalize and the general-equilibrium response of incumbent formal firms. The results underscore the importance of the competitive adjustment of large incumbents following the formalization of medium-sized informal firms.

5 Conclusion

This paper develops a general equilibrium model that integrates firm heterogeneity and market power within a framework where firms choose to operate either formally or informally. The model provides a framework to understand how the interaction between market power and informality shapes macroeconomic outcomes. We calibrate the model with rich Brazilian administrative data and perform several counterfactuals.

Reducing informality leads to gains in productivity and government revenue, although market power remains an important factor in shaping aggregate outcomes. These findings underline the importance of considering both market structure and informality in the design of economic policies in developing countries.

[14] In the counterfactual, no firm switches from formal to informal.

References

Atkeson, A. and A. Burstein (2008): "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 98, 1998–2031.

Bao, R., J. D. Loecker, and J. Eeckhout (2022): "Are Managers Paid for Market Power?" NBER Working Papers 29918, National Bureau of Economic Research, Inc.

Boar, C. and V. Midrigan (2023): "Markups and Inequality," Tech. rep., mimeo.

Burstein, A., V. M. Carvalho, and B. Grassi (2023): "Bottom-up Markup Fluctuations," Tech. rep., mimeo.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins," *American Economic Review: Papers & Proceedings*, 101, 471–475.

De Loecker, J. and J. Eeckhout (2021): "Global Market Power," Tech. rep., mimeo.

De Loecker, J., J. Eeckhout, and S. Mongey (2022): "Quantifying Market Power," Tech. rep., mimeo.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The Rise of Market Power and the Macroeconomic Implications*," *The Quarterly Journal of Economics*, 135, 561–644.

Edmond, C., V. Midrigan, and D. Y. Xu (2015): "Competition, Markups, and the Gains from International Trade," *American Economic Review*, 105, 3183–3221.

——— (2023): "How Costly Are Markups?" *Journal of Political Economy*, 131, 1619–1675.

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

Meghir, C., R. Narita, and J.-M. Robin (2015): "Wages and Informality in Developing Countries," *American Economic Review*, 105, 1509–46.

Ulyssea, G. (2018): "Firms, Informality, and Development: Theory and Evidence from Brazil," *American Economic Review*, 108, 2015–2047.

——— (2020): "Informality: Causes and Consequences for Development," *Annual Review of Economics*, 12, 525–546.

Online Appendix

A Additional Figures and Tables

Figure A.1

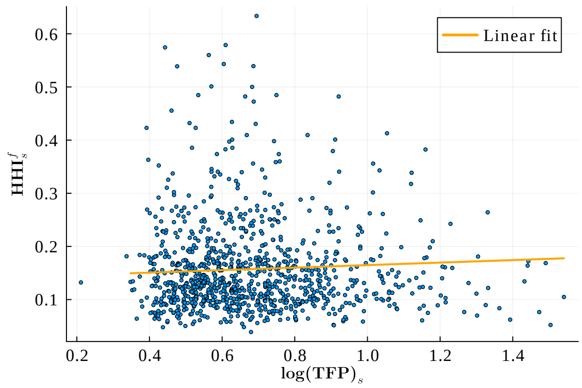


Table A.1: Decomposition of Ω in the low informality scenario

	Ω	μ
Baseline	1.028	1.235
No Informality	1.025	1.231
No Dispersion Within	1.007	1.238
No Dispersion Within + No Inf.	1.004	1.235
No Dispersion Within and Between	1.003	1.239
No Dispersion Within and Between + No Inf.	1.000	1.235

Notes:

B Model with Labor Only

TBW

C Model with Multiple Inputs

In this section, we extend the baseline model by including capital.

Firm's problem. Consider a constant return to scale production function with capital: $y_i = z_i k_i^\alpha n_i^{1-\alpha}$. Standard cost minimization implies that the marginal cost of firm i is

$$mc_i = \frac{1}{z_i}\left(\frac{r}{\alpha}\right)^\alpha \left(\frac{w}{1-\alpha}\right)^{1-\alpha} \tau_i^{1-\alpha} = \frac{MC}{z_i}\tau_i^{1-\alpha}, \tag{C.1}$$

where r is the rental rate of capital, $MC \equiv \left(\frac{r}{\alpha}\right)^\alpha \left(\frac{w}{1-\alpha}\right)^{1-\alpha}$ the aggregate marginal cost, and $\tau_i = 1 + \tau_w$ for a formal firm and $\tau_i = \tau_1 z_i^{\tau_2}$ for an informal firm.

It is straightforward to show that the production function with multiple inputs maps into a production function with only labor:

$$\begin{aligned}
y_i =& z_i k_i^\alpha n_i^{1-\alpha} = z_i \left(\frac{k_i}{n_i}\right)^\alpha n_i \\
=& \frac{w\tau_i}{mc_i}\frac{1}{1-\alpha} n_i = \underbrace{\frac{w}{MC}\frac{1}{1-\alpha} z_i \tau_i^\alpha}_{\equiv \tilde{z}_i} n_i \\
=& \tilde{z}_i n_i
\end{aligned}$$

where $z_i \left(\frac{k_i}{n_i}\right)^\alpha = \frac{w\tau_i}{mc_i}\frac{1}{1-\alpha}$ comes from the first order condition of the cost-minimization problem. Note that $\tilde{z}_i$ is different for formal and informal firms. Similarly, the total variable cost is:

$$\begin{aligned}
rk_i + \tau_i w n_i = mc_i y_i =& \frac{MC}{z_i}\tau_i^{1-\alpha} y_i \\
=& \underbrace{\frac{w}{1-\alpha}}_{\equiv \tilde{w}} \underbrace{\frac{1-\alpha}{w}\frac{MC}{z_i \tau_i^\alpha}}_{\equiv 1/\tilde{z}_i} y_i \tau_i \\
=& \frac{\tilde{w}}{\tilde{z}_i} y_i \tau_i = \tau_i \tilde{w} n_i.
\end{aligned}$$

Thus, by redefining z_i and w, the problem with multiple inputs is equivalent to the problem with only labor.

Finally, the problem of the formal firm is:

$$\pi^F_{is} = \max_{y_{is},p_{is}} \quad (1-\tau_y)p_{is}y_{is} - \frac{MC}{z_{is}}(1+\tau_w)^{1-\alpha}y_{is} - wc_F \tag{C.2}$$
$$\text{s.t.} \quad p_{is} = y_{is}^{-\frac{1}{\gamma}} y_s^{\frac{1}{\gamma}-\frac{1}{\theta}} Y^{\frac{1}{\theta}} P,$$

while the problem of the informal firm is:

$$\pi^I_{is} = \max_{y_{is},p_{is}} \quad p_{is}y_{is} - \frac{MC}{z_{is}}(\tau_1 z_{is}^{\tau_2})^{1-\alpha}y_{is} \tag{C.3}$$
$$\text{s.t.} \quad p_{is} = y_{is}^{-\frac{1}{\gamma}} y_s^{\frac{1}{\gamma}-\frac{1}{\theta}} Y^{\frac{1}{\theta}} P.$$

The solutions imply the usual pricing equations:

$$p^F_{is} = \mu_{is}\frac{MC}{z_{is}}\frac{(1+\tau_w)^{1-\alpha}}{1-\tau_y} \qquad \text{and} \qquad p^I_{is} = \mu_{is}\frac{MC}{z_{is}^{1-\tau_2(1-\alpha)}}\tau_1^{1-\alpha}.$$

Aggregation. Define a sectoral markup μ_s such that it satisfies:

$$p_s = \mu_s \frac{MC}{Z_s}\frac{(1+\tau_w)^{(1-\alpha)}}{(1-\tau_y)}, \quad \text{where} \quad Z_s \equiv \left[\sum_{i=1}^{M_s} z_{is}^{\gamma-1}\right]^{\frac{1}{\gamma-1}}$$

is the market-efficient level of productivity given the measure of firms M_s.

Define the market-level aggregate production function as

$$Y_s = \Omega_s^{-1} Z_s (K_s)^{\alpha} (N_s^{\text{Oper.}})^{1-\alpha}, \tag{C.4}$$

where $N_s^{\text{Oper.}}$ is the total labor in market s used in production (i.e., labor net of fixed cost), K_s is the total capital in market s, and Ω_s^{-1} the wedge reflecting misallocation in the market (if > 1). Note that aggregate capital and labor are given by:

$$K_s = \sum_i k_i = \sum_i mc_i \frac{\alpha y_i}{r} = \frac{\alpha MC}{r}\sum_i \frac{y_i}{z_i}\tau_i^{1-\alpha},$$
$$N_s^{\text{Oper.}} = \sum_i n_i = \sum_i mc_i \frac{(1-\alpha)y_i}{w\tau_i} = \frac{(1-\alpha)MC}{w}\sum_i \frac{y_i}{z_i}\tau_i^{-\alpha}.$$

Substituting in the aggregate production function:

$$
\begin{aligned}
Y_s =& \Omega_s^{-1} Z_s \left[\frac{\alpha MC}{r} \sum_i \frac{y_i}{z_i} \tau_i^{1-\alpha}\right]^{\alpha} \left[\frac{(1-\alpha) MC}{w} \sum_i \frac{y_i}{z_i} \tau_i^{-\alpha}\right]^{1-\alpha}, \\
\Omega_s Y_s =& Z_s \left[\sum_i \frac{y_i}{z_i} \tau_i^{1-\alpha}\right]^{\alpha} \left[\sum_i \frac{y_i}{z_i} \tau_i^{-\alpha}\right]^{1-\alpha}, \\
\Omega_s =& Z_s \left[\sum_i \frac{1}{z_i} \left(\frac{p_i}{p_s}\right)^{-\gamma} \tau_i^{1-\alpha}\right]^{\alpha} \left[\sum_i \frac{1}{z_i} \left(\frac{p_i}{p_s}\right)^{-\gamma} \tau_i^{-\alpha}\right]^{1-\alpha},
\end{aligned}
$$

where in the last line we use the fact that $y_i = \left(\frac{p_i}{p_s}\right)^{-\gamma} Y_s$. Substituting the definition of p_i/p_s and separating formal and informal firms:

$$
\begin{aligned}
\Omega_s =& \left[\underbrace{\sum_i \left(\frac{z_i}{Z_s}\right)^{\gamma-1} \tau_i^{1-\alpha} \left(\frac{\phi_i^F \mu_i^F + (1-\phi_i^F)\mu_i^I (\tau_1 z_i^{\tau_2})^{(1-\alpha)}(1-\tau_y)/(1+\tau_w)^{(1-\alpha)}}{\mu_s}\right)^{-\gamma}}_{\equiv \Omega_s^K}\right]^{\alpha} \dots \\
& \dots \times \left[\underbrace{\sum_i \left(\frac{z_i}{Z_s}\right)^{\gamma-1} \tau_i^{-\alpha} \left(\frac{\phi_i^F \mu_i^F + (1-\phi_i^F)\mu_i^I (\tau_1 z_i^{\tau_2})^{(1-\alpha)}(1-\tau_y)/(1+\tau_w)^{(1-\alpha)}}{\mu_s}\right)^{-\gamma}}_{\equiv \Omega_s^N}\right]^{1-\alpha}
\end{aligned}
$$

where $\tau_i = \phi_i^F(1+\tau_w) + (1-\phi_i^F)\tau_1 z_i^{\tau_2}$. The extra term τ_i increases the misallocation wedge, Ω_s, as it interacts with the informality decision and the dispersion of markups. This term arises because the labor tax and the informality cost distort the input decision (in our case, the capital-labor ratio). The wedge collapses to the one derived from the baseline model if (i) the model only has one input (i.e., $\alpha = 0$); or (ii) τ_i is constant across all firms.

D Computational Algorithm

TBW