Learning in the Shadows:
Informality and Entrepreneurship in Brazil

Yanran Guo [1*], Roberto Lagos Mondragon [2]

[1]York University

[2]University of North Carolina at Chapel Hill

Abstract

We examine the role of the informal sector in shaping entrepreneurial dynamics. Using Brazilian data, we document two novel empirical facts. First, around one-third of high-income entrepreneurs operate their businesses in the informal sector, and they closely resemble their formal sector counterparts across a range of characteristics. Second, high-income entrepreneurs are more likely to transition into the formal sector over time. These observations raise a central question: Why do these highly productive individuals choose to start out informally and only later formalize? To interpret these findings, we develop a quantitative model featuring imperfect information and learning. Individuals choose between wage employment and entrepreneurship without fully knowing their business potential. Within this framework, the informal sector endogenously arises as a cost-effective platform for entrepreneurial experimentation. Individuals operate informally to gradually learn about their business quality. Entrepreneurs who discover they are highly productive subsequently transition into the formal sector to expand production and access financial markets. The calibrated model replicates the observed transition patterns from informality to formality and generates policy counterfactuals consistent with historical reforms in Brazil. Specifically, the model shows that reducing entry costs alone has limited effects on formalization. In contrast, combining entry-cost reductions with temporary tax relief leads to substantially larger declines in informality. Importantly, the resulting increase in formal-sector firms is driven primarily by the formalization of existing informal businesses rather than by the creation of new formal firms.

Keywords: Informality, Entrepreneurship, Learning, Formalization policy, Occupation misallocation

JEL Codes: E26, O17, L26, D80, H20

---

[*]Corresponding author: yrguo@yorku.ca

We are grateful to Tasso Adamopoulos, Chaoran Chen, Lutz Hendricks, Gueorgui Kambourov, Pete Klenow, William F. Maloney, Stephen Ross, Katsuya Takii, Eugene Tan, Can Tian, and all seminar participants at the University of North Carolina at Chapel Hill, the University of Toronto, World Bank, York University, the University of Connecticut, the University of Osaka, the 2024 SEA Annual Meeting, XXII RIDGE FORUM: Growth and Development in Macroeconomics, and 2025 CIGS End of Year Macroeconomics Conference for their valuable questions, comments, and insights. We also thank Ying Feng for helpful discussions on the algorithm.

1. Introduction

Most developing countries have a large informal sector, contributing 30–70% of GDP and involving 20–80% of the labor force, as well as a significant share of enterprises (Ulyssea (2020)). The informal sector encompasses a wide range of economic activities, from street vending and unregistered ride-sharing to domestic work and handicrafts. These activities typically operate without formal business registration or labor contracts. As a result, a substantial portion of economic activity takes place outside the reach of tax, labor, and other regulatory frameworks. The prevalence of informality has significant implications for individual behavior and aggregate outcomes, affecting productivity, total output, and long run growth. Understanding the drivers and consequences of informality is thus essential for both economic research and policy design.

Using data from Brazil's National Household Sample Survey (PNAD, PNADC) and the Survey of the Urban Informal Economy (ECINF), which together provide detailed information on both formal and informal sector households and businesses, we document two empirical patterns that are new to the literature. First, over one-third of entrepreneurs in the top income decile operate in the informal sector, and they closely resemble their formal-sector counterparts across a range of characteristics, including skill level and firm productivity. Second, high income informal sector entrepreneurs are more likely to transition to the formal sector. These findings raise a central question: why do these highly productive individuals choose to begin their ventures informally and only later transition to formality?

To interpret these patterns, we develop a quantitative model of occupational and sectoral choice featuring imperfect information and learning. Individuals choose between wage employment and entrepreneurship, and, if the latter, whether to operate in the formal or informal sector. Sector choice applies only to entrepreneurs, meaning that workers supply labor to firms and earn the same wage rate regardless of the firm's registration status. We focus on the extensive margin of informality only: entrepreneurs choose to enter the formal or informal sector.[1]

The key feature of our framework is information friction. Individuals do not observe their business quality ex ante and must learn through experience. Entrepreneurs use production outcomes as signals to update their beliefs over time. Moreover, in each period, informal entrepreneurs have the option to transition into the formal sector, allowing the model to capture the rich dynamics observed in the data. Entering the formal sector requires paying a fixed registration cost and complying with ongoing tax obligations, while informal businesses avoid these costs but face stricter credit constraints and the risk of detection. Within this environment, the informal sector emerges endogenously as a platform for entrepreneurial experimentation. It offers a low-cost setting in which individuals can assess the viability of their business ideas. Those who learn that their firms are highly productive eventually choose to formalize in order to access the credit market and expand their operations without the threat of enforcement.

[1] Ulyssea (2018) distinguishes between the extensive margin (whether firms formally register) and the intensive margin (whether formal firms hire labor off the book). Since our empirical focus is on entrepreneurship transitions from informal to formal status, we abstract from the intensive margin.

Although not directly targeted by the parameterization, this learning-based mechanism generates transition patterns from informal to formal entrepreneurship that align closely with empirical evidence. In particular, the model replicates the observed 10% transition rate from informal to formal entrepreneurship, and captures the fact that this rate increases with income. We also contrast this benchmark model with a perfect information version in which individuals observe their business quality from the outset. This version fails to replicate the observed transition patterns, highlighting the role of information frictions and learning in shaping sectoral dynamics.

Our model thus sheds light on the role of informality as a stepping stone and a strategic decision for some entrepreneurs. Moreover, our framework emphasizes how the interaction between learning and sectoral selection affects the allocation of talent and resources in the economy. On one hand, informality facilitates entrepreneurial discovery by allowing households to experiment with uncertain business opportunities. This channel is particularly important in developing economies, where formal support systems and access to capital are limited. On the other hand, informality imposes meaningful constraints, including stricter borrowing limits and exposure to regulatory enforcement, which can impede long-run growth and reduce aggregate efficiency. Our model captures the trade-offs between these forces and provides a unified framework to study their implications for entrepreneurship, productivity, and aggregate outcomes.

We then use the calibrated model to evaluate the effects of different policies. While registration costs are often viewed as key barriers to formal sector entry (e.g., Djankov et al. (2002)), empirical studies find limited support for the effectiveness of entry cost reductions alone (La Porta and Shleifer (2014)). Recent evidence instead highlights the importance of reducing ongoing costs of formality or increasing the benefits of being formal. For instance, Rocha et al. (2018) study a large-scale formalization program in Brazil and find that reducing entry costs alone does little to lower informality, whereas combining entry cost reductions with tax relief yields substantially greater effects. Importantly, this impact is primarily driven by the formalization of existing informal firms, not the creation of new formal ones. Our framework provides a structural explanation for this empirical pattern.

Broadly following the structure of the reform studied in Rocha et al. (2018), we examine two policy counterfactuals. The first reduces the cost of entering the formal sector by 50%. The second builds on this by also eliminating the social security contribution for formal firms less than three years old, which corresponds to a 20-percentage-point reduction in payroll taxes. In line with empirical findings, we show that reducing entry costs alone has modest effects: the share of informal entrepreneurs among all entrepreneurs declines by 3.38 percentage points, the informal sector's share of entrepreneurial output falls by just 1.72 percentage points, and total entrepreneurial output increases by only 0.22%.

This limited effect reflects the composition of informal entrepreneurs in the model. Those who have completed the learning process and discovered they have low productivity do not find formalization worthwhile, even with lower entry costs, as their businesses would not survive under formal-sector obligations. High-productivity entrepreneurs who are still in the early stages of learning also prefer to remain informal, as it allows for lower-cost experimentation. As a result, the policy only affects entrepreneurs who are already near the margin of formalization, leading to modest aggregate impacts.

The second policy, which combines entry cost reductions with payroll tax relief for young formal firms, produces much larger effects. The share of informal entrepreneurs falls from 74% in the baseline economy to 60%, a reduction of 14 percentage points. The informal sector's share of entrepreneurial output shrinks by nearly half, and the number of formal-sector firms increases by 48%. Crucially, consistent with the data, this expansion is driven not by a larger inflow of new formal firms but by greater formalization among existing informal entrepreneurs. The number of entrepreneurs transitioning from informal to formal rises by 22%.

The mechanism behind this result lies in the reduced cost of learning in the formal sector. The temporary tax relief encourages entrepreneurs to shift part of the experimentation process into the formal sector. In the benchmark economy, it takes an average of 3.62 years for an entrepreneur to formalize. Under the reform, this duration shortens to 2.14 years. Because production in the formal sector remains costly, just less so than before due to the payroll tax relief, entrepreneurs still find it optimal to begin informally, using this phase to obtain a rough initial signal about their business quality. They transition only after gaining some confidence in their prospects. Thus, entrepreneurs do not immediately enter the formal sector but instead formalize until early uncertainty is resolved. The reform does not eliminate the need for learning but instead shifts part of the learning process into the formal sector. Therefore, this policy primarily affects the timing and trajectory of formalization, not the initial sector of entry.

This dynamic would not arise in a model with perfect information. In that setting, entrepreneurs know their business productivity from the outset. Those with high productivity but facing binding borrowing constraints are initially unable to enter the formal sector. Due to financial frictions, these constrained entrepreneurs rely more heavily on labor than on capital and are thus more exposed to payroll taxation. Such individuals would choose to start their businesses formally in the new steady state with a payroll tax relief. As a result, new formal entrants would play a larger role in driving the increase in formal-sector firms in the perfect information model.

This policy counterfactual also generates broader economic gains through improved talent allocation. Lower formalization costs and tax incentives accelerate the transition of entrepreneurs with high business productivity into the formal sector. In the benchmark economy, 55% of the most productive businesses operate formally; under the reform, this share rises to nearly 60%. These entrepreneurs benefit from formality by gaining better access to capital and expanding production without the risk of detection by authorities. As a result, total output produced by high-productivity businesses increases by 18%. In addition, measured entrepreneurial productivity rises by 1%, and overall entrepreneurial output grows by more than 7% in the new steady state.

Lastly, we analyze the effects of a policy that increases the cost of informal business operations, which is interpreted as intensified government monitoring. By explicitly modeling business learning, we highlight the informal sector's role as a stepping stone for entrepreneurship and show that raising the cost of informality can lead to unintended consequences. While the policy successfully reduces informality by cutting the number of informal entrepreneurs by nearly 40% and lowering the informal sector's share of output by 15 percentage points, it also imposes costs on the aggregate economy. Stricter enforcement raises barriers to entrepreneurial experimentation, resulting in a 16% decline in total entrepreneurship. By restricting informal entry, the policy weakens entrepreneurial selection. Consequently, entrepreneurial output falls by 3%, and aggregate output declines by more than 1%. These adverse effects stem from the disruption of the learning process: in the absence of a low-cost experimentation channel, many potential entrepreneurs remain in wage employment, and many of the most productive firms never emerge.

The rest of the paper is organized as follows. Section 2 frames our contribution in the context of the previous literature. Section 3 presents motivating facts on informality. Section 4 lays out the quantitative model and defines the equilibrium. Section 5 discusses the calibration procedure and evaluates the fit of the model against a number of important features of the data. The core of our analysis, the policy experiments, is presented in Section 6. Section 7 concludes.

2. Related Literature

This paper contributes to several strands of literature at the intersection of informality, entrepreneurship, and learning. It is motivated by new empirical evidence and develops a quantitative framework to interpret these facts and provide a structural interpretation of policy evaluation results. By explicitly modeling business learning and dynamic sectoral choice, the paper offers new insights into informality and formalization that complement and extend existing work.

A large literature studies the role of informality in developing economies. As summarized in Ulyssea (2018), there are three dominant views. The *survival view* interprets informality as a subsistence activity undertaken by low productivity individuals who are unable to operate formally. This perspective is closely related to work on subsistence entrepreneurship and self employment in developing countries (see, for example, Jaar (2026), Feng et al. (2024), and Herreño and Ocampo (2023)). The *parasite view* emphasizes that firms with sufficiently high productivity choose to remain informal in order to evade taxes or regulations, despite being able to survive in the formal sector. The *De Soto view* instead argues that the informal sector is composed of potentially productive entrepreneurs who are excluded from formality by high regulatory costs, particularly entry regulation. While these frameworks differ in their interpretation of informality, they largely adopt a static perspective in which firms are permanently formal or informal, especially in the survival and parasite views.

Our paper takes a different perspective by emphasizing the dynamic role of informality as a stepping stone for entrepreneurship. This perspective is motivated by two empirical patterns documented in the data. First, a nontrivial share of high income entrepreneurs operate in the informal sector, a fact that has been noted but not emphasized in previous work. Second, and novel to this paper, high income informal entrepreneurs exhibit a higher probability of transitioning into the formal sector over time. Taken together, these patterns are difficult to reconcile with static views of informality as a permanent state associated with low productivity or exclusion. Instead, they point to a dynamic sectoral choice in which informality serves as a temporary operating mode for productive entrepreneurs early in the business life cycle.

In our framework, entrepreneurs experiment and update their beliefs about business quality over time. Those who learn that their productivity is high eventually transition to the formal sector, while those who learn that their productivity is low remain informal as subsistence entrepreneurs. The main innovation of this paper is to shift the focus of the informality literature toward high income entrepreneurs, a group that has received relatively little attention in existing work. By studying informal sector entrepreneurs in the upper tail of the income distribution, we uncover a stepping stone role of informality that is largely absent from previous analyses and offer a new interpretation of informal activity as a deliberate, temporary choice rather than a permanent state.

We are not the first to argue that informality can serve as a stepping stone for entrepreneurship. A small but growing literature has begun to explore this perspective. Franjo et al. (2022) and Erosa et al. (2023) show that the informal sector can play a positive role in economies with limited financial access. Our paper builds on this insight but goes beyond financial frictions by incorporating business learning. As we demonstrate in our quantitative framework, financial constraints alone are insufficient to generate the magnitude of informal to formal transitions observed in the data. Accounting for learning about business quality is essential to match these transition dynamics.

Our analysis further contributes to the literature on misallocation and productivity. A large body of work documents the negative effects of misallocation on total factor productivity and aggregate output (e.g., Hsieh and Klenow (2009); Restuccia and Rogerson (2008); Guner et al. (2008); Garicano et al. (2016)). While this literature has convincingly established the aggregate consequences of misallocation, the underlying sources and mechanisms remain an active area of research. We contribute to this discussion by highlighting business learning and information frictions as an underexplored source of misallocation. In our model, uncertainty about business quality delays formalization and discourages entry by high productivity entrepreneurs, leading to inefficient allocation of talent and capital across sectors and occupations.

Building on the seminal work of Jovanovic (1982), a growing literature studies the role of learning in shaping firm dynamics and aggregate productivity (e.g., David et al. (2016); Chen et al. (2023)). Recent work by Feng (2025) provides empirical evidence of firm learning in developing countries. While this literature has emphasized the implications of learning for firm growth, exit, and productivity dynamics, much less is known about the quantitative importance of information frictions for occupational selection into entrepreneurship and sectoral choice. Gao and Zhang (2024) is among the first to explore this interaction. Our paper builds on this emerging strand by studying how business learning interacts with occupational and sectoral choices in an environment where informality is prevalent.

Our model also contributes to the literature on the effectiveness of formalization policies. A number of empirical studies evaluate reforms aimed at reducing entry costs or increasing the benefits of formality (e.g., Kaplan et al. (2011); Bruhn and McKenzie (2014)). Recent evidence finds limited support for the effectiveness of reducing entry costs alone as a tool for formalization (see, for example, La Porta and Shleifer (2014)). In particular, Rocha et al. (2018) show that cutting entry costs has only modest effects on informality, whereas combining entry cost reductions with lower tax burdens leads to substantially greater formalization. Despite these empirical findings, there is limited theoretical work that explains the mechanisms behind them.

Our paper addresses this gap. The quantitative model we develop aligns with the empirical evidence and provides a conceptual framework for interpreting it. By explicitly modeling entrepreneurial learning and sectoral transitions, we show that the effectiveness of formalization policies depends critically on how they alter the cost and incentives of experimentation faced by entrepreneurs.

Finally, our paper connects to a broader literature that examines policies aimed at fostering the growth of small and medium sized enterprises in developing countries. This literature studies how interventions such as tax cuts, capital subsidies, and reductions in labor costs can promote firm growth and economic development (e.g., De Mel et al. (2008); Karlan and Zinman (2010)). Despite its relevance, this line of work remains largely disconnected from the literature on informality. Bridging these two areas is an important and underexplored avenue for research. Our paper takes a step in this direction by showing that targeted subsidies for formal sector firms can also function as a formalization policy. We examine how such policies interact with firm level learning and shape entrepreneurial dynamics, formalization decisions, and aggregate economic outcomes.

3. Motivating Facts About Informality in Brazil

This section documents key empirical patterns related to entrepreneurs and firms in Brazil's formal and informal sectors, which serves as the motivation for our model. These facts also inform the calibration of the quantitative framework. We first describe the primary data sources and define the key concepts, and then present a set of stylized facts.

3.1. Data and Definitions

Our empirical analysis draws on microdata from Brazil, a country with both a large informal sector and rich survey data covering informal and formal business activity. The primary data sources are the PNAD (Pesquisa Nacional por Amostra de Domicílios) and the ECINF (Pesquisa de Economia Informal Urbana). In addition, we use the PNADC, a panel version of PNAD, to compute transition matrices. This dataset is discussed in detail in Section 3.4. All datasets are collected by the Brazilian Bureau of Statistics (IBGE).

PNAD is a nationally representative household survey that provides detailed information on labor market outcomes, income, occupation, formality status, and sociodemographic characteristics. It enables us to identify both formal and informal entrepreneurs and to examine their demographic profiles and basic firm characteristics, including age, gender, education, income, and firm size. However, as a household survey, PNAD focuses primarily on individual-level information. To supplement this, we use ECINF, a nationally representative survey of small urban businesses, which offers richer firm-level data. ECINF includes information on revenue, capital, employment, intermediate inputs, total costs, and other operational characteristics. While PNAD's primary objective is to capture labor market outcomes and socioeconomic characteristics, ECINF was specifically designed to measure the scale and features of informal economic activity in Brazil. IBGE implements rigorous methodological procedures to ensure the accuracy, reliability, and confidentiality of responses in both surveys. As a result, PNAD and ECINF have been widely used in research on informality due to their robust representation of informal sector activities.[2]

[2]These surveys rely on self-reported information, which naturally raises concerns about measurement error. However, IBGE has a long-standing reputation for accurately capturing informality and enforces strict confidentiality protocols, ensuring that responses cannot be used for auditing purposes. Moreover, we observe high levels of informality in these datasets, even among high-income individuals, which further supports the credibility of the data and suggests that respondents are not systematically underreporting their informality status.

Throughout our analysis, we identify entrepreneurs in the household survey data based on individuals' reported occupations. Specifically, individuals classified as self-employed or employers are considered entrepreneurs. This definition is applied consistently across both the empirical analysis and the theoretical model. The IBGE defines firm informality according to registration status. In particular, a firm is considered informal if it is not registered with the tax authority under a CNPJ (Cadastro Nacional da Pessoa Jurídica), Brazil's national business registration number. Unregistered firms are not subject to formal tax obligations, including contributions to the social security system (INSS—Instituto Nacional do Seguro Social), either on behalf of the owner or their employees if they hire external labor. These contributions are mandatory under Brazilian law and form part of the broader "encargos sociais" (social charges) within the payroll tax system. Following this definition, we identify informality using CNPJ registration when such information is available. In the PNAD data, we use reported contributions to social security system as a proxy for formality status.

Given that the most recent ECINF data available is from 2003, we restrict the PNAD sample to the years 2002–2009 to ensure comparability across datasets.[3] We also maintain consistent sample selection criteria across datasets whenever possible. Our analysis is limited to non-agricultural sector working-age individuals (18–65 years old). Unemployed individuals are excluded from the sample. This selection yields approximately 140,000 observations per survey year. About 95% of respondents report holding only one job, and we focus on information related to their primary occupation.[4]

3.2. Entrepreneurs in the Informal and Formal Sectors

Brazil exhibits both a high rate of entrepreneurship and a significant degree of informality among entrepreneurs, a pattern consistent with observations in many other developing countries. Entrepreneurs account for approximately 25% of the working population in Brazil, yet only 26% of them operate businesses in the formal sector.

A large body of literature using microdata from multiple countries documents that individuals in the informal sector tend to have lower skill levels, often measured by educational attainment, and that informal firms typically employ fewer workers and exhibit lower value added per worker compared to formal firms (La Porta and Shleifer, 2014). Our analysis of Brazilian data confirms these patterns. The majority of entrepreneurs, especially those in the informal sector, are self-employed individuals who run their businesses without hiring external labor. Among all self-employed individuals, 82% operate in the informal sector. Fewer than 30% of informal entrepreneurs have completed at least a high school education, in contrast to over 60% of formal entrepreneurs. Additionally, informal entrepreneurs earn significantly less than their formal sector counterparts. Despite these significant differences, formal and informal entrepreneurs often coexist within narrowly defined industries. This finding aligns with the results of Ulyssea (2020).

Building on this established evidence, our study takes a step further by exploring potential heterogeneity among informal entrepreneurs and their firms, differences that may be obscured in

[3] While PNAD is a repeated cross-sectional survey with different respondents each year, the key patterns we analyze remain stable over time. We therefore pool all available years.

[4] See Appendix for further details on data and measurement.

aggregate statistics. To uncover these patterns, we classify individuals into income deciles and analyze the characteristics of entrepreneurs and their firms in different income groups.

Figure 1 presents the population share of entrepreneurs across income deciles, distinguishing between formal (blue) and informal (red) entrepreneurs.[5] The distribution shows that formal entrepreneurs are largely absent from the lower half of the income distribution. Instead, they are concentrated at the upper end: 71% of formal sector entrepreneurs fall within the top three income deciles, and 37% are in the top decile alone.

Importantly, a large share of high-income entrepreneurs operate informally. Specifically, 37% of entrepreneurs in the top income decile run informal businesses, and 51% of those in the top three deciles are informal. This finding contributes a new perspective to the literature, which has traditionally emphasized that informal entrepreneurs tend to earn lower profits and operate less productive firms. While informality is indeed prevalent among lower-income entrepreneurs, our results indicate that it is not exclusive to the bottom of the distribution. A meaningful share of high-income entrepreneurs choose to stay informal. This raises important questions: Who are these high-income informal entrepreneurs, and what do their firms look like? How do they compare to their formal sector counterparts?

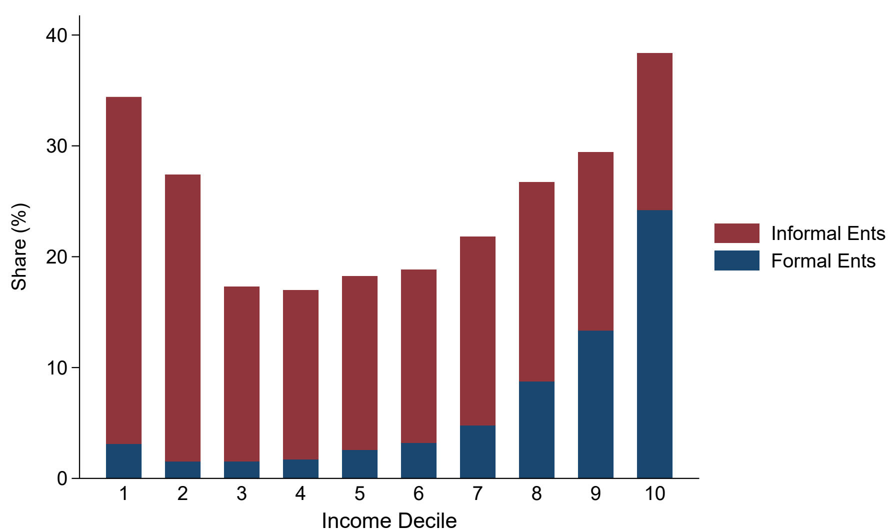

Figure 1. Share of Entrepreneurs in Each Income Decile

Notes: Data from PNAD. This figure illustrates the population share of entrepreneurs within each income decile. The blue segments represent formal entrepreneurs, while the red segments correspond to informal entrepreneurs.

[5]The income distribution is generated using income from all sources. A potential concern is that some individuals, especially those in the top decile, may appear to have high incomes primarily due to the large capital they are holding instead of their labor market activity. However, this does not appear to be the case in Brazil. On average, 94% of total income comes from individuals' main job. Even among those in the top income decile, more than 87% of total income comes from their main job.

3.3. High Income Informal Entrepreneurs

To address this question, we compare informal and formal entrepreneurs within the top income decile along four key dimensions: education, firm size, firm productivity, and business income.

For this analysis, we draw on both the PNAD and ECINF datasets. ECINF provides detailed information on firm characteristics, particularly with regard to productivity. A key challenge in using ECINF is that it is not a nationally representative household survey; rather, it is designed as a survey targeting non-agricultural businesses. To identify top decile entrepreneurs in ECINF, we rely on PNAD to estimate an income threshold. Based on PNAD data (in 2018 values), the cutoff for the top income decile is approximately R$4,000 per month. Additionally, we show that nearly 90% of income for individuals in the top decile comes from their main job. This implies that an entrepreneur would need to earn at least R$3,600 per month in business profits (R$4,000 $\times$ 0.9) to be considered part of the top income decile. We therefore classify ECINF entrepreneurs with monthly profits at or above R$3,600 as belonging to the top decile.

The PNAD and ECINF datasets appear broadly consistent in terms of both formality rates and income distributions. In PNAD, 26% of entrepreneurs are classified as formal; in ECINF, this figure is 24%. Table 1 reports the mean and median monthly profits for all entrepreneurs, as well as for formal and informal entrepreneurs separately, in both datasets. The income measures are closely aligned, reinforcing their comparability. Furthermore, the share of top-decile entrepreneurs is similar: 15% in PNAD and 16% in ECINF, based on our profit threshold. Among these high-income entrepreneurs, 37% are informal in PNAD, compared to 42% in ECINF. These patterns support the validity of our classification strategy and confirm the comparability of the two datasets.

R$ / month	PNAD		ECINF	
	Median	Mean	Median	Mean
All ents	1,067	2,165	1,155	2,271
Informal ents	811	1,389	924	1,600
Formal ents	2,667	4,377	2,680	4,467

Table 1. Median and Mean Profit: PNAD vs ECINF

Notes: This table reports the mean and median monthly profit from PNAD and ECINF, with all values expressed in 2018 Brazilian reais.

Figure 2 presents the distribution of formal (blue) and informal (red) entrepreneurs across four education categories: less than middle school; completed middle school but not high school; completed high school but not college; and college or above. The left panel shows this distribution for all entrepreneurs. Consistent with prior research, informal entrepreneurs tend to have lower levels of education. On average, entrepreneurs have 7.74 years of schooling, with informal and formal entrepreneurs averaging 6.89 and 10.19 years, respectively. More than half of informal entrepreneurs did not complete middle school, and only about 30% have attained a high school education or higher. In contrast, over 60% of formal entrepreneurs have completed at least high school, twice the share observed among informal entrepreneurs.

However, the right panel, restricted to entrepreneurs in the top income decile, reveals a different pattern. Among high-income entrepreneurs, educational differences between formal and informal sectors largely disappear. Nearly 70% of informal entrepreneurs in this group have completed high school or more, compared to 78% among formal entrepreneurs. This suggests that, in terms of human capital, high-income informal entrepreneurs more closely resemble their formal sector counterparts than the broader population of informal entrepreneurs.

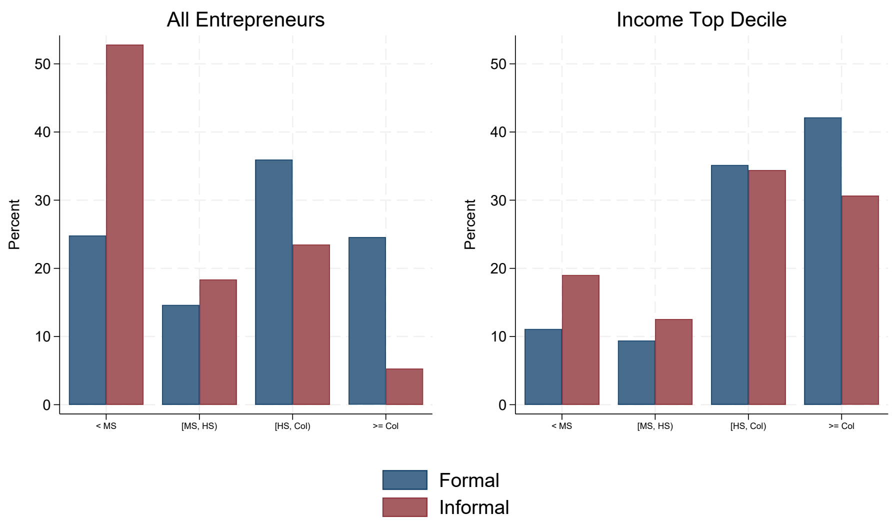

Figure 2. Entrepreneurs' Education Level

Notes: Data from PNAD. This figure shows the share of informal entrepreneurs (red) and formal entrepreneurs (blue) across different education levels. Education is categorized into four groups: less than middle school, completed middle school but less than high school, completed high school but less than college, and college or above.

Figure 3 presents the distribution of firm size for formal and informal entrepreneurs. Overall, self-employed individuals are heavily concentrated in the informal sector. Approximately 91% of informal entrepreneurs operate without any employees. In contrast, firms run by formal entrepreneurs tend to be larger. This size gap persists even among high income entrepreneurs, though it narrows somewhat. Among entrepreneurs in the top income decile, around 60% of informal business owners still operate without employees. One possible explanation for this persistent size difference is the strategic behavior of informal firms. Operating informally entails the risk of detection and penalties by regulatory authorities. To remain under the radar, many informal entrepreneurs may deliberately limit their firm size to avoid attracting attention from the government. Nonetheless, it is worth noting that large informal firms, those with more than five employees, still account for a non-negligible share: nearly 10% of top income decile informal entrepreneurs fall into this category.

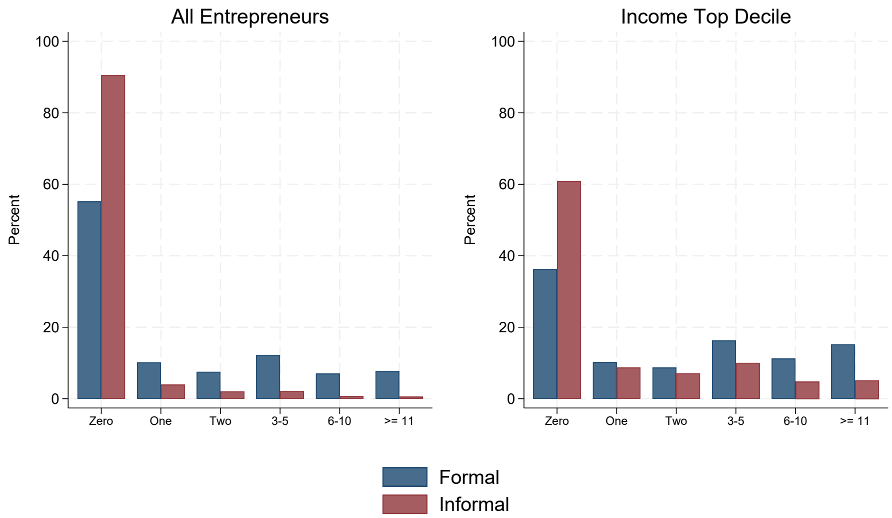

Figure 3. Firm Size Distribution

Notes: Data from PNAD. This figure shows the share of informal entrepreneurs (red) and formal entrepreneurs (blue) across different firm size categories.

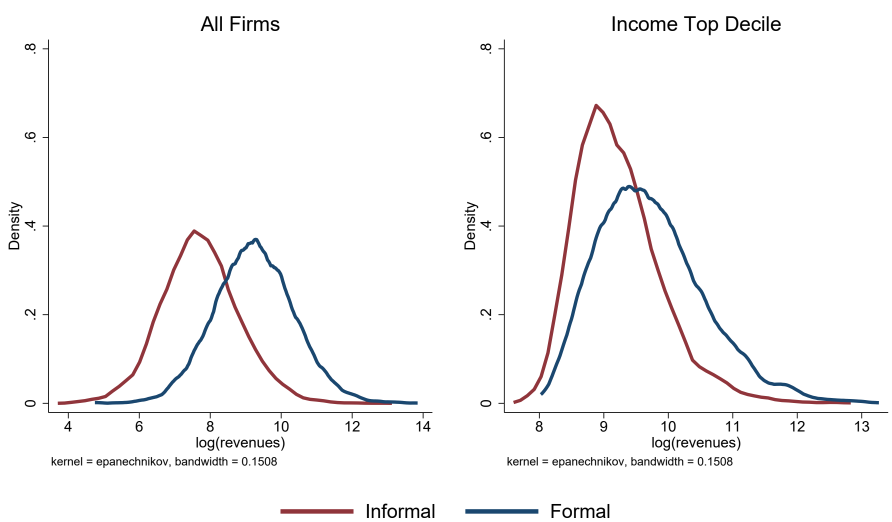

Figure 4. Firm Size Distribution (Revenue Measure)

Notes: Data from ECINF. We regress the log of revenues on a set of industry dummies. This figure shows the densities of computed log residuals for formal (blue) and informal (red) firms as a proxy for firm size. The left panel displays the size distribution for all firms, while the right panel focuses on firms operated by entrepreneurs in the top income decile.

To better capture the economic scale of firms, we also construct a revenue-based measure of firm size. Specifically, we regress the log of firm revenues on a set of industry dummies to control for inter-industry variation, and use the resulting log residuals as a proxy for firm size. Figure 4 displays the distribution of this measure for formal and informal firms. The left panel includes all firms, while the right panel is restricted to those owned by top income decile entrepreneurs. On average, the size distribution for formal firms is substantially shifted to the right, indicating that they tend to operate at a larger scale. However, among high-income entrepreneurs, there is considerable overlap in the size distribution of formal and informal firms, suggesting that some informal businesses may reach a comparable economic scale.

Next, we examine productivity differences between formal and informal firms. As a proxy for firm level productivity, we regress the log of value-added per worker on industry fixed effects to remove variation attributable to inter-industry variation. Figure 5 displays the resulting productivity distributions for informal (red) and formal (blue) firms. The left panel, which includes all firms, shows that the productivity distribution for formal firms is clearly shifted to the right, indicating that, on average, formal firms are more productive than informal firms. However, the right panel, which restricts the sample to entrepreneurs in the top income decile, reveals substantial overlap between the productivity distributions of formal and informal firms. This suggests that many high-income informal firms operate at productivity levels comparable to those of formal firms, reinforcing the idea that informality at the top of the income distribution is not necessarily associated with low firm quality.

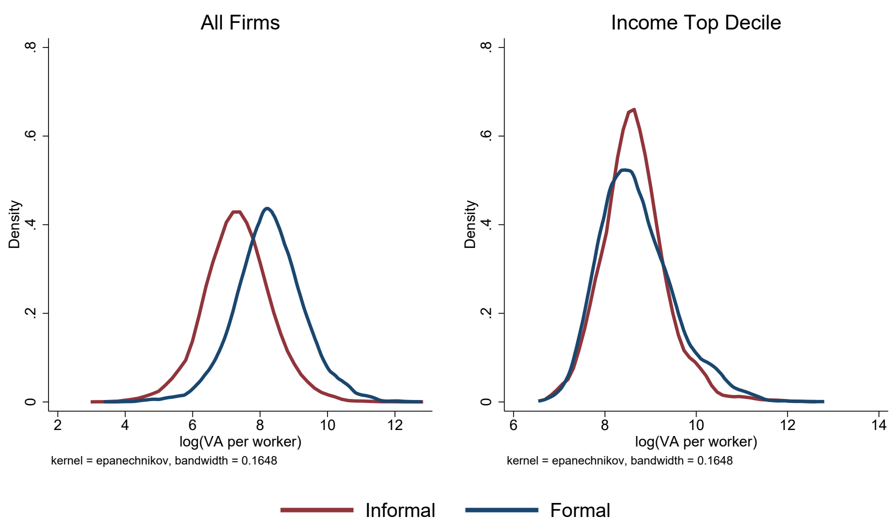

Figure 5. Firm Productivity Distribution

Notes: Data from ECINF. We regress the log of value-added per worker controlling for industry. The figure presents the density distributions of the computed log-residuals for formal and informal firms. The left panel displays the productivity distribution for all firms, while the right panel focuses on firms operated by entrepreneurs in the top income decile.

Finally, we examine the profit gap between formal and informal entrepreneurs. The previous literature documents that individuals in the informal sector earn less, on average, than those in the formal sector. Our analysis confirms this pattern within the entrepreneurial population. After controlling for individual characteristics, such as education, gender, age, and race, as well as business characteristics including location, industry, size, and firm age, we find that formal entrepreneurs earn, on average, 40% more than their informal counterparts (see Column 1 of Table 2).[6] However, this income differential is not uniform across the distribution. When we restrict the sample to entrepreneurs in the top income decile, the gap narrows substantially to less than 4% (see Column 2 of Table 2). A similar pattern is observed in the ECINF data, as shown in Columns 3 and 4.

These findings suggest that, while informal entrepreneurs on average tend to have lower education levels, operate smaller and less productive firms, and earn lower profits, those in the top income decile differ markedly from this general pattern. In terms of education, productivity, and profit, high-income informal entrepreneurs closely resemble their formal sector counterparts. This challenges the view that informality is purely a necessity-driven outcome. Instead, it points to a more nuanced picture: for at least a subset of entrepreneurs, informality may reflect a strategic choice rather than a binding constraint.

	PNAD		ECINF	
	(1)	(2)	(3)	(4)
$log(profit)$	All Ents	Top Decile	All Ents	Top Decile
Formal ent	0.337***	0.036***	0.389***	0.094***
Male	0.491***	0.131***	0.430***	0.074**
Age	0.050***	0.006***	0.060***	-0.003
Age squared	-0.000***	0.000	-0.001***	0.000
Firm age	0.009***	0.003***	0.337***	0.035***
Educ level	Yes	Yes	Yes	Yes
Race	Yes	Yes	No	No
State	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
Firm size category	Yes	Yes	Yes	Yes
Observations	282,544	41,302	21,356	4,165
R-squared	0.519	0.191	0.447	0.151

*** p<0.01, ** p<0.05, * p<0.1

Table 2. Income Gap Between Formal and Informal Entrepreneurs

Notes: Data from PNAD and ECINF. The table reports the profit gap between formal and informal entrepreneurs when controlling for both individual and firm characteristics.

[6]This corresponds to an estimated coefficient of 0.337 in a log income regression, which translates to a 40% difference: $100 \times (\exp(0.337) - 1) = 40\%$.

3.4. Transitions from Informal to Formal Entrepreneurship

To examine whether informality reflects a strategic choice for some entrepreneurs, we analyze transitions between informal and formal entrepreneurship. This analysis provides deeper insight into the mechanisms underlying informality. If entrepreneurs use the informal sector as a temporary stage, we should observe transitioning into the formal sector as their businesses grow and succeed. In particular, we would expect many high-income informal entrepreneurs to formalize over time. Conversely, if high-income informal entrepreneurs consistently remain informal, this would suggest that other factors, such as tax avoidance, regulatory burdens, or institutional frictions, are key barriers to formalization. Such a pattern would challenge the stepping stone hypothesis and indicate that even successful entrepreneurs may be locked into informality due to structural disincentives.

To conduct the transition analysis, we use data from the Continuous PNAD (PNADC), a rotating panel version of the PNAD survey. PNADC tracks individuals over five consecutive quarters, allowing us to observe transitions between the informal and formal sectors over a one-year period. The survey was launched by IBGE in 2012, but the 2012–2013 waves are regarded as a transitional series due to the gradual rollout and methodological adjustments made during those initial years. Moreover, Brazil experienced a recession between mid-2014 and mid-2016, officially emerging from it in the first quarter of 2017. To avoid periods marked by either data quality concerns or economic instability, we base our transition analysis on the 2017 wave of PNADC.[7] To ensure that observed transitions reflect changes in the informality status of the same business, rather than exits and entries into entirely new businesses, we examine firm-level characteristics across quarters. This helps confirm that transitions between informal and formal status occur within continuing businesses.

		t		
$t-1$	Worker	Informal Ent	Formal Ent	Total
Informal Ent (All)	12.34	77.58	10.08	100
Informal Ent (Decile 1-4)	14.12	80.18	5.70	100
Informal Ent (Decile 8-10)	10.04	68.77	21.18	100
Informal Ent (Top decile)	11.07	61.92	27.00	100

Table 3. Transition From Informal to Formal Entrepreneurship

Notes: Data from PNADC. The table reports the one-year transition rates of informal entrepreneurs into three categories: wage workers, informal entrepreneurs, and formal entrepreneurs. The first row presents transition rates for all informal entrepreneurs. The second row focuses on informal entrepreneurs in the lower-income group (income deciles 1–4), while the last two row report transitions for high income informal entrepreneurs: income deciles 8–10 and the top income decile.

[7] Similar results are obtained using the 2018 wave.

Table 3 reports one year transition rates of informal entrepreneurs into three categories: wage employment, continued informal entrepreneurship, and formal entrepreneurship. The first row presents results for all informal entrepreneurs. On average, 10.08% transition into the formal sector after one year, while the majority, nearly 80%, remain in the informal sector. Transition rates vary by income level. Among low income informal entrepreneurs (defined as those in the bottom four income deciles), only 5.70% formalize within a year (second row). In contrast, informal entrepreneurs in the upper income deciles are more likely to transition. Over 21% of those in income deciles 8 to 10 transition into the formal sector, and the rate rises to 27% for those in the top income decile (third and fourth rows, respectively). These patterns suggest that informality is not always a long-term condition, but may instead reflect a temporary and strategic choice for some entrepreneurs.

To further explore the determinants of transitions from informal to formal entrepreneurship, we run a regression where the dependent variable is a dummy variable which indicates whether an informal entrepreneur formalizes their business within one year. We control for individual characteristics (income, education, gender, age) and firm level indicators (firm size, recent firm growth, location, and industry). Table 4 reports the estimation results. We find that higher income, higher educational attainment, larger firm size, and recent expansion in employment are significantly associated with a higher likelihood of transitioning to the formal sector.

$\mathbf{1}_{In \to F}$	
log(income)	0.035***
Education	0.006***
Male	0.012**
Age	0.005***
Age squared	-0.000***
Firm size	0.030***
Growth	0.185***
Race	Yes
State	Yes
Industry	Yes
Observations	12,726
R-squared	0.108

*** p<0.01, ** p<0.05, * p<0.1

Table 4. Transition From Informal to Formal Entrepreneurs

Notes: Data from PNADC. The table reports the transition from informal to formal entrepreneurship when controlling for both individual and firm characteristics.

4. Model

In the empirical section, we document two novel facts. First, more than one third of entrepreneurs in the top income decile operate in the informal sector, and these high income informal entrepreneurs closely resemble their formal-sector counterparts along key observable dimensions. Second, high income informal entrepreneurs are more likely to transition to the formal sector. To interpret these findings, we develop a quantitative model in which individuals make occupational and sectoral choices under incomplete information about their business potential. The model incorporates business learning and allows us to analyze how informality may serve as a platform for experimentation. This framework provides a lens through which to understand the dynamic relationship between informality, learning, and formalization decisions.

4.1. Environment

Time is discrete, and each period corresponds to one year. The economy consists of a government, a representative corporate firm, and a continuum of heterogeneous agents. In each period, agents choose whether to work as wage earners or become entrepreneurs, and, if the latter, whether to operate in the formal or informal sector. Thus, the model focuses on the extensive margin of informality on the firm side. In particular, entrepreneurs decide between formal and informal operation, while workers do not choose sectors. All workers supply labor inelastically and receive the same wage rate, though their labor income may vary due to heterogeneity in individual working productivity.

Following the informality literature (e.g., Ulyssea (2018)), the distinction between the formal and informal sector arises from differences in tax enforcement and the probability of detection by government authorities. Informal firms avoid paying taxes but, in doing so, face frictions that limit their ability to grow, which is modeled as a labor distortion that prevents them from hiring large workforces.[8] Entrepreneurs can enter the formal sector by paying a fixed entry cost. In addition, we incorporate collateral constraints following Buera and Shin (2013), such that the capital used by entrepreneurs is limited by their asset holdings. Importantly, informal sector entrepreneurs face tighter financial constraints, reflecting their limited access to formal credit markets. As a result, informality is associated not only with regulatory distortion but also with restricted access to external financing.

A central innovation of our model is the incorporation of business learning. We assume that households do not observe their true business quality when they make decisions. Instead, entrepreneurs learn about their underlying productivity over time by observing realized business outcomes and updating their beliefs. This learning friction plays a central role in shaping occupational dynamics and the strategic use of informality as a platform for experimentation.

[8] Ulyssea (2018) show that this labor distortion setup can be specialized to a formulation that explicitly accounts for a detection probability.

4.2. Demographics and Preferences

Building on Erosa (2001) and Franjo et al. (2022), we adopt an overlapping generations model. As illustrated in Figure 6, agents enter the economy at the age of 20 ($t = 1$), and make occupational decisions between wage employment and entrepreneurship in each period throughout their working years. Upon reaching age 65 ($t = 46$), agents are mandatorily retired. From that period on, each agent faces an age-dependent probability of death (Ω_t). For simplicity, we assume a zero mortality rate before age 65, reflecting the small share of adults in Brazil who die prior to this age. This assumption has minimal impact on the results. Agents face a mortality rate of 1 at age 75, meaning they die with certainty by this age. When agents die, they are replaced by their descendants at age 20, inheriting all accumulated wealth. There is no bequest motive, meaning that bequests are accidental. Furthermore, the model operates in a stationary environment, and variables are indexed only by model age (t), with the time index omitted for clarity.

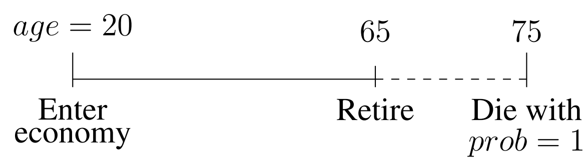

Figure 6. Life Cycle of the Model

Preferences are time-separable, with a constant discount factor, β. The utility derived from consumption in each period is represented by the following function:

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma} \tag{1}$$

where c denotes consumption, and γ is the coefficient of relative risk aversion, capturing the curvature of the utility function.

4.3. Endowments

Agents are endowed with one unit of time each period. In addition, each agent possesses two types of skills: one that determines their productivity as a worker and another that governs their productivity in entrepreneurial activities. As a result, agents make occupation choices at each period during their working years, based on these two productivity as well as other state variables.

Labor Market Productivity: Labor market productivity for individual i at age t, denoted by ω_{it}, is specified as follows:

$$\log(\omega_{it}) = \log(g(t)) + \log(s_{it}) \tag{2}$$

where $g(t)$ is an exogenous age-dependent component that is common to all agents at the same age, and s_{it} is a persistent idiosyncratic working productivity shock that follows an AR(1) process:

$$\log(s_t) = \rho_s \log(s_{t-1}) + \epsilon_{st}, \quad \epsilon_{st} \sim \mathcal{N}(0, \sigma_s^2) \tag{3}$$

Business Productivity: At the start of each period, working-age individuals who are not yet entrepreneurs draw a business idea with an innate quality q from a normal distribution with mean μ_q and variance σ_q^2, i.e., $q \sim \mathcal{N}(\mu_q, \sigma_q^2)$. If the agent chooses to become an entrepreneur, the drawn q becomes fixed and remains constant. That is, the business's innate productivity is determined at entry and does not evolve thereafter. If the agent does not choose entrepreneurship in that period, the business idea is discarded, and a new idea is drawn in the following period. A key feature of the model is that agents do not observe the value of q at the time of decision-making. As a result, agents make their occupation and sectoral choice without knowing the true quality of their business idea.

The overall business productivity in each period includes two components: the fixed innate business quality and a transitory shock:

$$Q_t = q + e_t \tag{4}$$

where e_t represents an $i.i.d$ transitory shock drawn from a normal distribution, $\mathcal{N}(0, \sigma_e^2)$, each period.

4.4. Technologies and Profits

Entrepreneurial Firms: All entrepreneurial firms, regardless of whether they operate in the informal or formal sectors, use the same decreasing returns to scale production function:

$$y = e^Q \Big(k^\alpha n^{1-\alpha}\Big)^\eta \tag{5}$$

where $\eta < 1$ is the span-of-control parameter. A fraction η of the output is allocated to factor inputs, with a share α going to capital and $1-\alpha$ allocated to labor. The production function also allows entrepreneurs to supply their own labor, ω, as an input to production, in addition to hiring external labor, $n-\omega$. Entrepreneurs rent capital at rate r and hire labor at rate w, choosing their capital and labor inputs in each period to maximize profits.

There are no entry costs for the informal sector. If an entrepreneur decides to operate in the informal sector, their profits are given by:

$$\pi^i = e^Q \Big(k^\alpha n^{1-\alpha}\Big)^\eta - (r+\delta)k - w(n-\omega) \cdot \mathbf{1}_{n>\omega} - w\tau(n) \tag{6}$$

where $r+\delta$ is the cost of renting capital (r is the rent rate and δ is the depreciation rate of capital), and $w(n-\omega)$ represents the wage bill for external labor. The indicator function $\mathbf{1}_{n>\omega}$ ensures that wages are only paid to external labor when $n > \omega$. Following Ulyssea (2018), we assume that informal sector entrepreneurs can avoid taxes but face a probability of detection by government officials. This expected cost takes the form of a labor distortion denoted by $\tau(n)$, which is assumed to be increasing and convex in firm size ($\tau' > 0$, $\tau'' > 0$). This assumption reflects the fact that larger firms face a higher probability of being detected, as larger firms are more visible to tax authorities, as suggested by De Paula and Scheinkman (2011).

If an entrepreneur decides to enter the formal sector, they must pay a one-time fixed entry cost, denoted by C^f_{entry}. The profit generated by a formal sector business is given by:

$$\pi^f = (1-\tau_y)e^Q \Big(k^\alpha n^{1-\alpha}\Big)^\eta - (r+\delta)k - w(1+\tau_{ss})(n-\omega) \cdot \mathbf{1}_{n>\omega} \tag{7}$$

Formal firms are registered with the tax authority, which means they must comply with all relevant taxes and regulations by paying the sales tax τ_y and payroll tax τ_{ss}.

Corporate Sector: In addition, we model a second production sector, referred to as the non-entrepreneurial corporate sector, which consists of a large number of homogeneous firms that are not directly managed by households. This sector is represented by a single representative corporate firm operating under a constant returns to scale production function:

$$\Pi_c = A_c K_c^{\alpha} N_c^{1-\alpha} - (r+\delta)K_c - wN_c \tag{8}$$

where A_c denotes the time-invariant corporate productivity, normalized to one without loss of generality. The variables K_c and N_c represent the corporate sector's demand for capital and labor, respectively. The outputs produced by the corporate and entrepreneurial firms are assumed to be perfect substitutes. Additionally, capital depreciates at a rate δ, and the corporate firm doesn't have any financial friction.

4.5. Learning

Agents do not observe their innate business quality, q. As a result, their occupation choice and, if they choose entrepreneurship, their production input decisions are based on their beliefs about q. Initially, agents have a prior belief about q that follows the population distribution of business quality, $\mathcal{N}(\mu_q, \sigma_q^2)$. When an agent decides to become an entrepreneur, whether in the formal or informal sector, they begin to learn about their innate business quality, q, through the process of production. Specifically, each period, entrepreneurs observe their input choices and output. From this, they can derive the overall business quality for that period, Q_t, but they cannot separately identify the fixed innate quality q and the transitory shock e_t.

Thus, entrepreneurs use Q_t as a noisy signal to update their beliefs about q using Bayesian inference. Initially, the agent's belief about q follows the prior distribution $\mathcal{N}(\hat{\mu}_{q,n}, \hat{\sigma}_{q,n}^2)$, where n denotes the number of periods the entrepreneur has been in operation, and equivalently, the number of signals they have observed. The update process for the entrepreneur's belief is then given by:

$$\begin{aligned} \hat{\mu}_{q,n+1} &= \frac{\hat{\sigma}_{q,n}^2 Q_t + \sigma_e^2 \hat{\mu}_{q,n}}{\hat{\sigma}_{q,t}^2 + \sigma_e^2} \\ \hat{\sigma}_{q,n+1}^2 &= \frac{\hat{\sigma}_{q,n}^2 \sigma_e^2}{\hat{\sigma}_{q,n}^2 + \sigma_e^2} = \frac{\sigma_q^2 \sigma_e^2}{(1+n)\sigma_q^2 + \sigma_e^2} \end{aligned} \tag{9}$$

In this process, $\hat{\mu}_{q,n+1}$ is the updated belief about q, and $\hat{\sigma}_{q,n+1}^2$ is the updated variance. The belief update reflects the prior information as well as the signal observed during the period. The uncertainty in the entrepreneur's belief about q decreases over time as more signals (or periods) are observed, which improves their ability to assess their business's true quality.

4.6. Asset Market and Borrowing Constraints

Agents have access to competitive financial intermediaries, who accept deposits from both workers and entrepreneurs and rent out capital to entrepreneurs and the corporate firm. Our analysis focuses on within-period borrowing, where capital can only be rented for production purposes.

This implies that agents cannot borrow for intertemporal consumption smoothing, resulting in a non-negativity constraint on financial wealth: $a \geq 0$. Following Franjo et al. (2022), we assume that access to credit requires compliance with business regulations, including registration with the tax authorities, which makes production activities observable to the government. As a result, only formal sector entrepreneurs can borrow from financial intermediaries, subject to a collateral constraint:

$$k \leq \lambda_f a \tag{10}$$

where $\lambda_f \geq 1$. In contrast, informal sector entrepreneurs rely solely on their own assets to finance capital input, facing the stricter constraint:

$$k \leq \lambda_i a \tag{11}$$

where $\lambda_i = 1$.

4.7. Government and Tax Systems

The government finances an exogenously determined level of public expenditure, denoted by $\overline{G}$, through taxes on individual consumption and on formal sector firms. Specifically, consumption is taxed at a flat rate τ_c. In addition, formal sector entrepreneurs are subject to a payroll tax at rate τ_{ss} on the wage payments made to hired labor, as well as a sales tax at rate τ_y levied on their total business revenue.

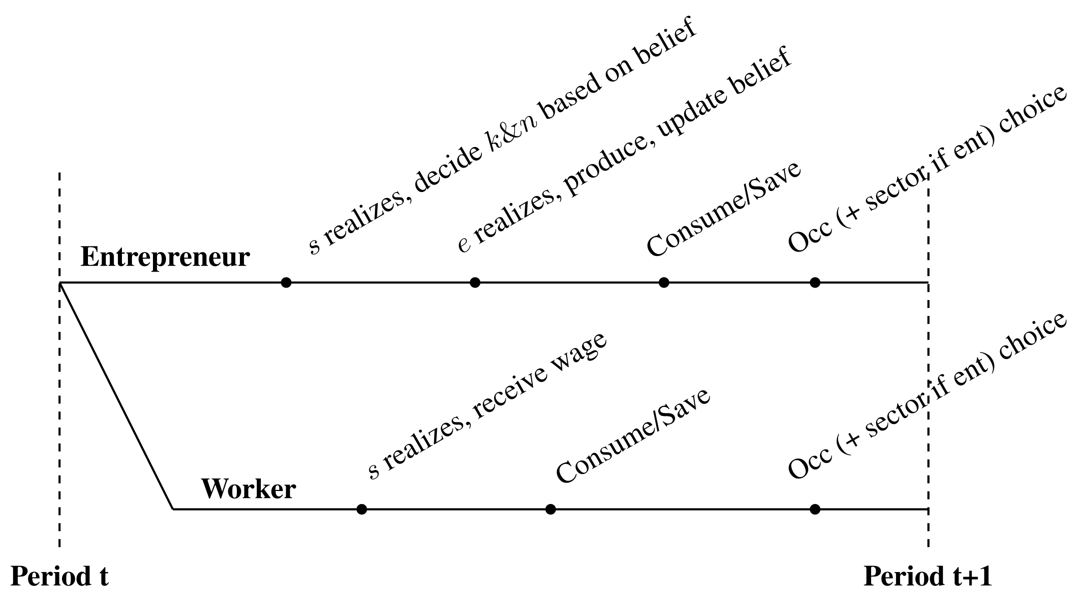

Figure 7. Timeline with One Period

4.8. The Agent's Recursive Problem

The precise timing of decisions and events within a period is summarized in Figure 7. We assume that occupation choices for period t is made at the end of period $t-1$, before observing the realization of labor market productivity s_t and without knowing the true innate business quality q. If an agent chooses entrepreneurship, they must also decide whether to enter the formal or informal sector. Entrepreneurs in the informal sector have the option to transition to the formal sector in the next period, whereas formal sector entrepreneurs cannot switch to informality. To simplify the analysis, we assume that only entrepreneurs make sectoral choices, and workers do not choose between sectors. Instead, labor is perfectly mobile across sectors, earning the same equilibrium wage w regardless of the sector in which they work.

Once occupation and sectoral choices are made, the economy moves to period t. At the beginning of the period, labor market productivity s_t is realized and becomes observable to all agents. Workers supply labor and earn wage income $w \cdot \omega_t$. Worker then make consumption and savings decisions before choosing their occupation for the next period. Given their own labor productivity, entrepreneurs decide optimal capital and labor input to maximize their period profit based on their belief about the innate business quality q. After these input choices are made, a transitory business productivity shock e_t is realized, determining the firm's output. Entrepreneurs do not directly observe the shock but can infer their overall business productivity Q_t using observed output and input levels. They then update their belief about q using Bayesian updating, incorporating both their prior belief and the noisy signal from Q_t. Finally, with the profits they earn in period t, entrepreneurs make consumption and savings decisions before determining their occupation for the next period.

In this model, the state variables for a working-age agent are given by $x = (a, s, Q, \hat{\mu}_q, n, t)$, where a is the agent's asset holdings, s is the current working productivity, Q is the overall business productivity inferred from production input and output. The variable $\hat{\mu}_q$ captures the agent's belief about its innate business quality q, while n represents the number of observed signals, which determines the variance of this belief. Finally, t denotes the agent's age. For retired agents, the state space simplifies to just age and asset holdings, given by (a, t).

Old Retirees: From $t = 46$ to $t = 55$ (corresponding to physical ages 65 to 74), agents are retired and no longer participate in the labor market. They survive to the next period with a probability that declines with age. Let Ω_t denote the age specific mortality rate. Hence, the probability of survival at age t is given by $1 - \Omega_t$. Old households consume and save on the basis of the financial wealth accumulated during their working life. Hence, the value function of retired household is given as

$$\begin{aligned} V^r(a,t) = \max_{c,a'} \; & u(c) + \beta(1-\Omega_{t+1})V^r(a', t+1) \\ \text{s.t.} \quad & (1-\tau_c)c + a' = (1+r)a \\ & a' \geq 0 \end{aligned} \tag{12}$$

Working Age Young: From $t = 1$ to 45 (physical age 20 to 64), households are in their working years and survive with certainty to the next period. During this phase, they choose between wage employment and entrepreneurship. If they opt for entrepreneurship, they must also decide whether to operate in the formal or informal sector. Once an entrepreneur enters the formal sector, they cannot revert their current business to the informal sector, whereas informal entrepreneurs have the option to formalize their business.

If an agent chooses to be a worker, they earn wage income by supplying their efficient labor, then make consumption and saving decisions before determining their occupation for the next period. The worker's value function is given by:

$$\begin{aligned}
V^w(a,s,t) = \max_{c,a'}\ & u(c) + \beta \cdot \mathbf{1}_{t=45} \cdot (1-\Omega_{t+1}) V^r(a', t+1) \\
+ \beta(1-\mathbf{1}_{t=45}) \max \Big\{ & \mathbb{E}[V^w(a', s', t+1)], \mathbb{E}[V^i(a', s', Q', \mu_q, 0, t+1)], \\
& \mathbb{E}[V^f(a' - C^f_{entry}, s', Q', \mu_q, 0, t+1)] \Big\} \\
\text{s.t.} \quad & (1-\tau_c)c + a' = w \cdot \omega + (1+r)a \\
& a' \geq 0
\end{aligned} \tag{13}$$

where C^f_{entry} is the one-time fixed entry cost required to start a formal sector business. The functions $V^i(\cdot)$ and $V^f(\cdot)$ represent the value of being an informal and formal sector entrepreneur, respectively.

Informal entrepreneurs do not pay any taxes for their production, but, in avoiding taxation, they face distortions in their production decisions and incur an additional cost, $\tau(n)$. The decision problem of an informal sector entrepreneur is formulated as follows:

$$\begin{aligned}
V^i(a,s,Q,\hat{\mu}_q,n,t) = \max_{k,n,c,a'}\ & u(c) + \beta \cdot \mathbf{1}_{t=45} \cdot (1-\Omega_{t+1}) V^r(a', t+1) \\
+ \beta(1-\mathbf{1}_{t=45}) \max \Big\{ & \mathbb{E}[V^w(a', s', t+1)], \mathbb{E}[V^i(a', s', Q', \hat{\mu}'_q, n+1, t+1)], \\
& \mathbb{E}[V^f(a' - C^f_{entry}, s', Q', \hat{\mu}'_q, n+1, t+1)] \Big\} \\
\text{s.t.} \quad & \pi^i = e^Q \Big(k^\alpha n^{1-\alpha}\Big)^\eta - (r+\delta)k - w(n-\omega) - w\tau(n) \\
& (1-\tau_c)c + a' = \pi^i + (1+r)a \\
& k \leq a, \quad a' \geq 0
\end{aligned} \tag{14}$$

The recursive problem of formal sector entrepreneurs closely resembles that of informal sector entrepreneurs, with several key distinctions. Formal entrepreneurs are subject to a payroll tax on hired labor and a sales tax on business revenue but do not face the production distortion that limits the scale of informal firms. Additionally, formal entrepreneurs have access to external financing, allowing them to borrow capital from financial intermediaries subject to collateral constraints. A further difference lies in the set of occupational choices. Formal entrepreneurs may continue operating their business or choose to shut it down and return to wage employment. However, once a business has entered the formal sector, it cannot revert to informality, that is, transitions from formal to informal entrepreneurship are not allowed.

$$
\begin{aligned}
V^f(a,s,Q,\hat{\mu}_q,n,t) = & \max_{k,n,c,a'} \; u(c) + \beta \cdot \mathbf{1}_{t=45} \cdot (1-\Omega_{t+1}) V^r(a',t+1) \\
& + \beta(1-\mathbf{1}_{t=45}) \max\Big\{\mathbb{E}[V^w(a',s',t+1)], \mathbb{E}[V^f(a',s',Q',\hat{\mu}'_q,n+1,t+1)]\Big\} \\
\text{s.t.} \quad & \pi^f = (1-\tau_y) e^Q \Big(k^\alpha n^{1-\alpha}\Big)^\eta - (r+\delta)k - (1+\tau_{ss}) w (n-\omega) \\
& (1-\tau_c)c + a' = \pi^f + (1+r)a \\
& k \le \lambda_f a, \quad a' \ge 0
\end{aligned} \tag{15}
$$

4.9. Stationary Competitive Equilibrium

Let us use $x = \{a, s, Q, \hat{\mu}_q, n, t\}$ to denote the state vector. And define $X = \{x, o(x)\}$ the vector containing the individual state variables and the occupational status $o(x)$, which can be retired, workers, informal sector entrepreneurs, and formal sector entrepreneurs. A stationary competitive equilibrium is given by a price vector $\{r, w\}$, allocations $\{c(X), a'(X)\}$, occupation choices $o(x)$, labor hired by formal and informal entrepreneurs $\{l^i(X), l^f(X)\}$, capital rented by formal and informal entrepreneurs $\{k^i(X), k^f(X)\}$, labor and capital demanded by the corporate firm $\{N_c, K_c\}$, and a distribution of agents over each state $\zeta(X)$, such that, given the prices and the tax system $\{\tau_c, \tau_y, \tau_{ss}\}$:[9]

1. Policy functions solve the agents' decision problems described in Sector 4.8.
2. The factors demand of the corporate sector solve equation 8.
3. The capital, labor, and goods markets clear. In particular

$$
\text{Capital market:} \int \Big(k^i(X) + k^f(X)\Big)\, d\zeta(X) + Kc = \int a(X)\, d\zeta(X) \tag{16}
$$

$$
\text{Labor market:} \int \Big(l^i(X) + l^f(X)\Big)\, d\zeta(X) + Nc = \int \omega_t \times 1_{\{o=W\}}\, d\zeta(X) \tag{17}
$$

   where $1_{\{o=W\}}$ is an indicator function taking value 1 if the agent is a worker and 0 otherwise.
4. The government budget is balanced

$$
\overline{G} = \int \Big\{\tau_c c(X) + 1_{\{o=F\}} \times \Big(\tau_y y(X) + \tau_{ss} w l(X)\Big)\Big\}\, d\zeta(X) \tag{18}
$$

   where $\overline{G}$ is the public expenditure, $y(X)$ is the entrepreneur's output. $1_{\{o=F\}}$ is an indicator function taking value 1 if the agent is a formal sector entrepreneur and 0 otherwise.
5. The distribution $\zeta(X)$ is the invariant distribution for the economy.

[9] $\{l^i(X), l^f(X)\}$ is the amount of outside labor hired by entrepreneurs to make up the difference between their optimal labor demand for production, n, and the amount of labor they can supply themselves, ω. That is, outside labor is given by $l = n - \omega$.

5. Quantitative Analysis

In this section, we describe the process of parameterizing the model in a stationary equilibrium. The model is calibrated through a two-step procedure using the simulated method of moments. In the first step, a subset of the parameters is externally calibrated, relying on estimates that are independent of the model or commonly used values in the literature. The parameters determined in this step are presented in Table 5. In the second step, the remaining six parameters are endogenously calibrated within the model to match key features of the Brazilian economy. The parameters calibrated to match model generated moments with those observed in the data are listed in Table 6.

5.1. Parameter Values Set Exogenously

Demographics: Within the model framework, one model period corresponds to one year. Agents enter the economy at the age of 20 (model age $t = 1$) and retire at the age of 65. During their working years, agents transition to the next period with certainty. Upon retirement at age 65, they face an age-dependent mortality risk, denoted by $\{\Omega_t\}_{t=46,\ldots,55}$, which increases with age. The mortality rates are calibrated using estimates from the complete life table published by IBGE[10]. By physical age 75 (model age $t = 56$), all agents exit the economy with certainty. This threshold is chosen to align with the average life expectancy in Brazil over the 2010–2018 period, as estimated by the World Bank.

Preferences and Bequest: The coefficient of risk aversion γ is set to be 1.5, consistent with the bulk of the literature on occupational choice models (e.g., Buera and Shin (2013)). This value is also in line with empirical estimates of the relative risk aversion coefficient for Brazil, which suggest a range of γ between 1 and 3 (Fajardo et al. (2012)). The discount factor is set at a standard level commonly used in the literature, specifically $\beta = 0.92$ (see, for example, Erosa et al. (2023)).

Working Productivity: The working productivity follows an AR(1) process with persistence ρ_s and unconditional variance σ_s^2. The parameter values are drawn from Conesa et al. (2009) where $\rho_s = 0.98$ and $\sigma_s = 0.17$. The working productivity process is then discretized following the methodology proposed by Tauchen (1986). In addition, we incorporate a deterministic age specific productivity profile, denoted by $g(t)$, which is estimated using earnings data from the PNAD.

Production Technologies: The capital share of production function α is set to be 0.406 as in Allub and Erosa (2019) and Franjo et al. (2022). The span of control parameter η is set to be 0.802, following Allub and Erosa (2019) and Franjo et al. (2022). This parameterization choice implies that the capital share in the entrepreneurial sector within my model is $\alpha\eta = 0.33$, a value that aligns closely with those employed in macroeconomic literature on entrepreneurship (see, for example, Buera et al. (2011)). We set the capital depreciation rate δ to be 0.05, which is a standard value in the literature, as in Guvenen et al. (2023).

[10]IBGE Demography and social statistics: `https://www.ibge.gov.br/en/statistics/full-list-statistics/17117-complete-life-tables.html?edicao=42019`

Government and Tax Systems: The government imposes a flat consumption tax at a rate of $\tau_c = 0.15$, as calibrated by Jung and Tran (2012) in their OLG model for the Brazilian economy. The sales tax rate is set at $\tau_y = 0.0925$, corresponding to the PIS/COFINS tax applied in Brazil. PIS/COFINS is a federal-level tax levied on the gross revenue of formal sector firms across nearly all sectors, including commerce, services, and industry. It is one of the most broadly paid federal taxes by formal firms in Brazil. The payroll tax rate is set at $\tau_{ss} = 0.29$, which includes both the employer's social security contribution (20 percent) and a direct payroll tax component (9 percent).

Parameter	Description	Source	Value
Preference			
γ	Relative risk aversion	Buera and Shin (2013)	1.50
β	Discount factor	Erosa et al. (2023)	0.92
Productivity			
ρ_s	Working prod. persistence	Conesa et al. (2009)	0.98
σ_s	Working prod. std dev	Conesa et al. (2009)	0.17
Production			
α	$e^Q(k^\alpha n^{1-\alpha})^\eta$	Allub and Erosa (2019)	0.406
η		Allub and Erosa (2019)	0.802
δ	Capital depreciation rate	Guvenen et al. (2023)	0.050
Taxes			
τ_c	Consumption tax	Jung and Tran (2012)	0.150
τ_y	Sales tax	PIS/COFINS	0.093
τ_{ss}	Payroll tax	S.S + Direct payroll tax	0.290

Table 5. Parameters Calibrated Outside of the Model

5.2. Parameter Values Calibrated Within the Model

There are six parameters that are internally calibrated within the model. First, informal sector firms face a labor distortion, denoted by $\tau(n)$, which is increasing and convex in firm size. Following Ulyssea (2018), we assume the functional form $\tau(n) = \frac{n^2}{b}$, where $b > 0$ is a parameter to be calibrated. Two additional parameters are related to formal sector entrepreneurship: the one-time fixed entry cost, denoted by C^f_{entry}, and the collateral constraint, represented by λ_f. The remaining three parameters govern the process for business quality. Specifically, an entrepreneur's overall productivity Q_t consists of two components: a fixed innate business quality q, drawn from a normal distribution $\mathcal{N}(\mu_q, \sigma_q^2)$ at the time of entry, and a transitory shock e_t, which is an i.i.d. draw from $\mathcal{N}(0, \sigma_e^2)$ in each period. The initial belief of startup entrepreneurs is also assumed to follow $\mathcal{N}(\mu_q, \sigma_q^2)$. Therefore, the parameters μ_q, σ_q^2, and σ_e^2 jointly govern the evolution of entrepreneurial productivity Q_t and the learning process.

All six parameters are calibrated jointly. In what follows, we provide intuition for how variation in the data, combined with the model's structure, identifies each parameter. We begin with μ_q, which represents the mean of the initial belief about innate business quality. Since households decide whether to enter entrepreneurship based on this belief, a higher μ_q increases the perceived expected return from entrepreneurship, thereby encouraging more households to start a business.

Holding other factors constant, a larger μ_q leads to a higher entrepreneurship rate in equilibrium. We therefore calibrate μ_q to match the entrepreneurship rate, defined as the share of entrepreneurs in population. This rate is estimated to be 25% based on PNAD data.

σ_q represents the standard deviation of the population distribution of the innate business quality q. Once a firm has an accurate belief about its q, this fixed component becomes the primary determinant of firm size, as e_t is a transitory shock and input decisions are made prior to its realization. The parameter σ_q determines the dispersion of firm sizes. A larger σ_q will lead to a thicker right tail in the firm size distribution. To calibrate σ_q, we target the share of formal firms with at least 51 employees, which captures the upper tail of the size distribution. Following Ulyssea (2018), we set this moment to 0.02, based on data from RAIS, an administrative dataset that covers the universe of formal sector firms in Brazil.

σ_e governs the dispersion of the transitory shock to business productivity. For entrepreneurs with accurate beliefs about their innate quality q, fluctuations in profits are primarily driven by this transitory component, since input decisions are made before the realization of e_t. If σ_e were zero, profit would be perfectly predictable from one period to the next; conversely, a larger σ_e implies greater volatility in profits across periods. We calibrate σ_e by targeting the one year autocorrelation of profits among formal sector entrepreneurs, which we estimate to be 0.74 using PNADC data. This identification strategy relies on the assumption that entrepreneurs use the informal sector as a low cost platform for learning about their business quality. As beliefs become more accurate and business turns out to be profitable, high productivity entrepreneurs are more likely to formalize. As a result, formal sector entrepreneurs are disproportionately those with precise beliefs about q, meaning that the residual variation in their profits primarily reflects the transitory shock e_t rather than learning noise.

Following Erosa et al. (2023), we calibrate the collateral constraint parameter λ_f to match the credit-to-output ratio among formal-sector entrepreneurs, which is estimated to be 0.43. The entry cost C^f_{entry} influences the share of formal businesses in the economy, as a higher entry cost discourages formal sector entry. We calibrate this parameter by targeting the fraction of formal businesses, which is estimated to be 0.26 using PNAD data. The final parameter, b, governs the severity of the labor distortion faced by informal firms. The distortion $\tau(n)$ is decreasing in b. A lower value of b implies a stronger penalty on firm size in the informal sector, discouraging the employment of hired labor. We calibrate b by targeting the share of informal firms with zero employees, which is estimated to be 0.91 in PNAD data.

5.3. Calibration Results

The model economy performs well in matching the targeted moments. Table 6 presents the calibration results, including parameter values, targeted moments, and comparisons between model and data moments.

The model successfully replicates the high rate of entrepreneurship observed in Brazil, where entrepreneurs make up 25% of the labor force, compared to the model's prediction of 26%. To generate this high entrepreneurship rate, the model requires the initial belief about innate business quality, μ_q, to be set at 0.06. Additionally, the model captures the right tail of the size distribution within the formal sector, although it slightly underestimates it. While the fraction of formal firms with at least 51 employees is 0.02 in the data, the model predicts 0.01. However, Figure 8 shows that the model fits the entire firm size distribution in the formal sector well, despite only using the right tail as a targeted moment for calibration. The calibration also closely matches the one-year profit autocorrelation of formal sector entrepreneurs by setting the standard deviation of the transitory shock to business productivity at 0.37.

The fixed entry cost to the formal sector in our model is calibrated to be 0.28, which corresponds to approximately R$4,873 in 2018 values. This estimate falls within the range reported in the existing literature. For example, the Hong Kong Trade Development Council (HKTDC) estimates the monetary cost of entering the formal sector in Brazil at around R$2,000, though this figure varies across regions.[11] However, entry into the formal sector also involves administrative procedures and paperwork, which can take anywhere from one month to over a year depending on the region.[12] The Inter-American Development Bank estimates that the effective cost of formalization in Brazil is around 74% of Gross National Income (GNI) per capita, or roughly R$6,300 in 2018 values. Our calibrated value lies within this range, between R$2,000 and R$6,300, and is close to the estimate reported by Ulyssea (2018), who finds an entry cost of approximately R$4,300. These values suggest that the entry cost is not prohibitively high. In our model, it amounts to less than 20% of average worker earnings, indicating a moderate barrier to formal sector entry.

Formal sector entrepreneurs can borrow up to 44% of their asset holdings, which implies a maximum investment capacity of $\lambda_f = 1.44$ times their own assets. This borrowing constraint generates a credit-to-output ratio in the formal sector of 0.43, consistent with the data. In addition, by setting the labor distortion parameter to $b = 3.19$, the model replicates the observed prevalence of single-person firms in the informal sector. In the model, the fraction of self-employed informal entrepreneurs is 0.92, compared to 0.91 in the data.

Parameter	Value	Targeted Moment	Data	Model
μ_q	0.059	Population share of ents (PNAD)	0.25	0.26
σ_q	0.686	Share of formal firms with $\geq$ 51 employees (RAIS)	0.02	0.01
σ_e	0.368	Formal ents' profit autocorrelation after 1 yr (PNADC)	0.74	0.74
λ_f	1.440	Credit/output formal ent (Erosa et al. (2023))	0.43	0.43
C^f_{entry}	0.279	Share of formal ents (PNAD)	0.26	0.26
b	3.188	Share of informal firms with 0 emp (PNAD)	0.91	0.92

Table 6. Parameters Calibrated (Jointly) Inside the Model

[11] Hong Kong Trade Development Council (HKTDC): `https://research.hktdc.com/en/article/MzgyMTg4ODM4`

[12] The Doing Business Project, World Bank: `www.doingbusiness.org`
For example, in the year 2010, it takes 16 procedures and 120 days on average to start a business in the formal sector.

We also assess how the model performs in comparison to selected non-targeted moments, with results reported in Table 7. The most important finding is that our model of business learning, entrepreneurship, and sectoral choice successfully reproduces these patterns despite not using them as calibration targets.

The model successfully replicates key empirical patterns related to the transition dynamics of high income informal entrepreneurs. In the data, the overall one year transition rate from informality to formality is 0.10. This rate rises to 0.21 among entrepreneurs in income deciles 8–10, and reaches 0.27 in the top income decile. The model generates similar patterns, with corresponding transition rates of 0.07, 0.19, and 0.24, respectively.

Non-targeted Moments	Data	Model
Share of informal ents who transition to formal	0.10	0.07
Share of informal ents in income deciles 8–10 who transition to formal	0.21	0.19
Share of top decile informal ents who transition to formal	0.27	0.24
Difference in the average age of formal & informal firms (years)	3	3.98
Difference in the average age of formal & informal ents (years)	3	3
Average years to transition from informal to formal	7	3.62
Exit rate of formal sector firms	0.13	0.18

Table 7. Model Performance: Untargeted Moments

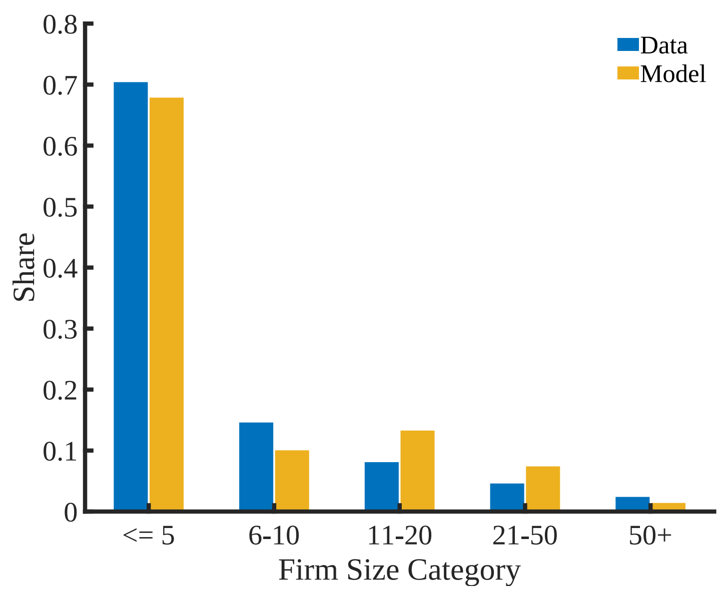

Figure 8. Firm Size Distribution in the Formal Sector

Notes: The firm size distribution data are taken from Ulyssea (2018), who estimate it using administrative records from the RAIS dataset.

In addition, the model captures age related differences across sectors. According to PNAD data, informal entrepreneurs are, on average, 40 years old, while their formal sector counterparts are 43. The model closely mirrors this pattern, generating average ages of 38 and 41, respectively. A similar pattern emerges for firm age: formal sector firms are, on average, three years older than informal firms in the data. The model produces a comparable age gap of approximately four years.

The model also performs well in replicating the exit rate of formal firms. Ulyssea (2020) estimates this rate at 0.13, while the model yields an exit rate of 0.18. Finally, we examine the distribution of firm size in the formal sector. Figure 8 compares the model generated distribution to estimates from Ulyssea (2018), based on administrative records from RAIS. Although our calibration explicitly targets only the upper tail, the model matches the entire firm size distribution well.

We then examine whether the model can replicate an important empirical pattern in the data: the decline in dispersion of marginal revenue products of capital (MRPK) with firm age. Following Feng (2025), we compute the dispersion of MRPK by firm age using ECINF data. Based on our production function, the MRPK of firm i in log terms is defined as:

$$mrpk_i = \log(\alpha\eta) + \log(y_i) - \log(k_i) \tag{19}$$

The y_i is measured as value-added, and k_i is measured as firm's capital stock. We calculate dispersion within each industry-age cell. Specifically, for each bin defined by firm age j and industry s, we require at least two firms and compute the standard deviation of $mrpk$ within that cell, denoted σ_{sj}. The weighted average dispersion for each firm age j is then:

$$\sigma_j = \sum_s \sigma_{sj} \cdot \omega_s, \quad \text{where } \omega_s = \frac{N_{sj}}{\sum_s N_{sj}} \tag{20}$$

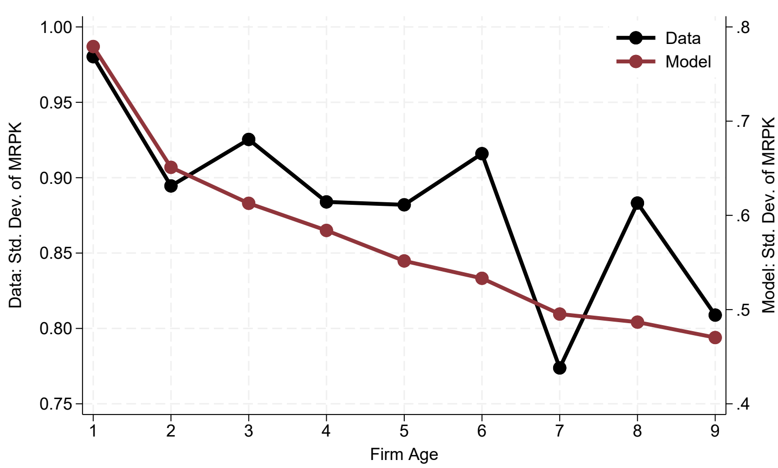

Figure 9. Dispersion of MRPK by Firm Age

Notes: Data from ECINF. The left y-axis corresponds to the empirical data, while the right y-axis corresponds to the model-simulated results.

Figure 9 plots σ_j by firm age. The black line corresponds to the data. Empirically, we observe that MRPK dispersion declines by nearly 0.2 points by firm age 9, a magnitude comparable to that documented by Feng (2025) using Chinese firm-level panel data. The reduction in MRPK dispersion over the firm life cycle provides indirect support for the model's learning mechanism: as firms age, they gain more information about their productivity and allocate capital more efficiently.

We acknowledge, however, that the observed pattern may also reflect compositional effects, particularly the gradual exit of less productive firms over time. Ideally, this would be addressed using a balanced panel, but ECINF is cross-sectional and does not allow us to track firm dynamics. To ensure consistency, we apply the same empirical procedure to the simulated data from our model, pooling firms by age without constructing a panel. Our model qualitatively replicates this declining trend in MRPK dispersion with firm age.

5.4. Recalibration Under Perfect Information

To better understand the role of information frictions and learning in shaping entrepreneurial behavior, we consider an alternative environment in which individuals have perfect information about their business quality. In this setting, agents observe the draw of their innate business quality q, prior to making occupational and sectoral choices. After choosing entrepreneurship, they still face transitory shocks, e, to productivity. Thus, the perfect information case is a nested version of our benchmark model, with the learning channel shut down. We recalibrate all six parameters, μ_q, σ_q, σ_e, λ_f, C^f_{entry}, and b, using the same set of empirical moments as in the benchmark model. This allows us to assess the consequences of shutting down the channel of information frictions while maintaining a consistent comparison with the data.

Panel A					
Parameter	Value (Info Fric)	Value (Perfect Info)	Targeted Moment	Data	PI Model
μ_q	0.059	0.071	Ent share	0.25	0.25
σ_q	0.686	0.368	Formal size ≥ 51	0.02	0.01
σ_e	0.368	0.220	Formal profit autocorr	0.74	0.74
λ_f	1.440	1.304	Formal credit/output	0.43	0.43
C^f_{entry}	0.279	9.554	Formal ent share	0.26	0.26
b	3.188	0.928	Informal se share	0.91	0.91
Panel B			Untargeted Moment	Data	PI Model
			Inf $\rightarrow$ F share	0.10	0.01
			Inc decile 8-10: Inf $\rightarrow$ F share	0.21	0.03
			Top inc decile: Inf $\rightarrow$ F share	0.27	0.06

Table 8. Parameters Calibrated (Jointly) in the Model with Perfect Information

Panel A of Table 8 presents the recalibrated parameter values. The second column reports values from the benchmark model with information frictions (henceforth, the IF model), while the third column shows the results from the perfect information model (PI model). The final column displays the model generated moments. Several key differences emerge between the IF and PI models, reflecting the important role that information frictions play in shaping equilibrium outcomes.

First, in the IF model, agents share a common prior and some choose to enter entrepreneurship not because they are inherently high quality entrepreneurs, but because they are uncertain about their potential. These "experimenting" entrants increase the overall entrepreneurship rate. In contrast, under perfect information, all agents observe their q directly and only those with sufficiently high productivity choose to become entrepreneurs. As a result, the margin of entry driven by uncertainty disappears. To match the same entrepreneurship rate as in the data, the mean of the business quality distribution, μ_q, must be higher in the PI model (0.071) than in the IF model (0.059).

Second, because high-q agents always enter entrepreneurship in the PI model, the standard deviation σ_q must be smaller. A higher σ_q would otherwise lead to an excessively fat upper tail in the firm size distribution. In the IF model, however, many high-q individuals may choose not to start a business due to uncertainty about their productivity. To compensate for this and match the empirical firm size distribution, the IF model requires a larger σ_q.

The calibrated value of σ_e also differs between the two models due to differences in the scale of q. In the benchmark model with information frictions, the greater dispersion in q implies that entrepreneurs operate with a wider range of underlying business productivity levels. To match the observed fluctuations in profits, the transitory shock must be sufficiently large to generate variation relative to the higher baseline productivity levels. In contrast, the perfect information model features a more compressed distribution of q, which reduces the scale of underlying productivity. Consequently, a smaller transitory shock is sufficient to generate comparable variation in observed profits. This structural difference leads to a higher calibrated value of σ_e in the benchmark model than in the perfect information model.

The most notable differences between the two models emerge in the calibration of the labor distortion parameter b and the formal sector entry cost C^f_{entry}. In the PI model, the estimated b is significantly lower, which is below 1 compared to 3.19 in the benchmark model. It implies a much stronger distortion in the informal sector in PI model. This result reflects the fact that in the IF model, many informal entrepreneurs do not hire workers simply because they are uncertain about their business quality, not necessarily because of the labor distortion. Once information frictions are removed, however, the only way to prevent informal firms from expanding is to impose a stronger distortion through a lower b.

Finally, without uncertainty about business quality, entrepreneurs have no incentive to remain in the informal sector for the purpose of experimentation. As a result, the model must rely more heavily on other frictions, namely, entry costs and financial constraints, to match the observed share of formal sector entrepreneurs. In particular, the calibrated model requires a tighter borrowing limit for formal sector firms: the collateral constraint parameter declines from 1.44 in the benchmark model to 1.30. To further limit excessive formal sector entry, the model also requires a higher entry cost. In the PI model, the calibrated entry cost is 9.55, approximately four times the average earnings of young workers.

Panel B of Table 8 presents key untargeted moments related to transitions from informal to formal entrepreneurship. The PI model fails to replicate the transition dynamics observed in the data. While the empirical transition rate from informal to formal entrepreneurship is 10%, the PI model predicts a rate of only 1%. It also significantly underestimates transition rates among high-income informal entrepreneurs: in the data, 27% of informal entrepreneurs in the top income decile transition to the formal secto, whereas the PI model generates a transition rate of just 6%. This result suggests that financial frictions and entry costs, while important, are not enough to account for the observed entrepreneurial and sectoral dynamics.

6. Policy Counterfactuals

We use the calibrated model to conduct two sets of policy counterfactuals. The first set is designed to provide a structural interpretation of the empirical findings in Rocha et al. (2018), who study a large scale formalization program in Brazil. They find that reducing entry costs alone has limited effects on informality, whereas combining entry cost reductions with lower tax rates leads to substantially larger declines in informality. Importantly, the expansion of the formal sector in their data is driven primarily by the formalization of existing informal firms rather than by the creation of new formal businesses. Broadly following the structure of this reform, we consider two counterfactual policies. The first reduces the cost of entry into the formal sector. The second builds on this policy by additionally offering temporary payroll tax relief to young formal sector firms. These experiments allow us to assess whether the model can replicate the observed patterns and to identify the mechanisms underlying them.

The second set of counterfactuals is designed to highlight the role of entrepreneurial learning in shaping policy outcomes. Specifically, we study a policy that increases the cost of informal operation, interpreted as intensified government monitoring and enforcement. In models where informality arises as an outcome underlying economic distortions, such policies typically improve aggregate outcomes by discouraging informal activity. In contrast, our framework highlights that informality can also serve a productive role by facilitating entrepreneurial experimentation and learning. As a result, stronger enforcement can reduce entrepreneurial entry and weaken selection, leading to lower aggregate output. This counterfactual illustrates how accounting for learning fundamentally alters the aggregate implications of informal sector enforcement.

In all policy experiments, we hold all other parameters fixed at their benchmark values, including the level of government expenditure $\overline{G}$. Any change in tax revenue resulting from policy reforms is offset by a lump-sum transfer T to households, ensuring that the government budget remains balanced. All counterfactuals are conducted in both partial equilibrium (PE) and general equilibrium (GE), and we focus exclusively on steady state comparisons.

6.1. Reducing Entry Costs

While registration costs are often cited as a key barrier to firm entry and formalization, recent evidence suggests that lowering these costs alone may have limited effects on informality (see, for example, La Porta and Shleifer (2014); Bruhn and McKenzie (2014)). Using Brazilian data, Rocha et al. (2018) document similar findings. In this subsection, we use the model to assess whether it can replicate these empirical patterns and to identify the mechanisms underlying the limited effectiveness of entry cost reductions.

Table 9 reports the results of a counterfactual policy that reduces the fixed entry cost to the formal sector by 50% (i.e., $C^f_{\text{entry}} \times \frac{1}{2}$). As expected, in partial equilibrium, lower entry costs lead to a modest decline in informality: the number of informal entrepreneurs decreases by approximately 2%, the share of informal entrepreneurs among all entrepreneurs falls from 74% to 70%, and the informal sector's share of total entrepreneurial output declines from 20% to 18%. The resulting expansion of the formal sector contributes to increases in output and tax revenue.

The limited impact of this policy stems from the underlying reasons why entrepreneurs choose to operate informally in our model. Due to information frictions, agents often enter the informal sector to experiment with their business ideas, as the cost of operating business in the informal sector is lower. Once they update their beliefs about business quality, their formalization decision depends on what they have learned. Entrepreneurs who discover they have low productivity choose to remain informal, as they cannot cover the ongoing costs of operating formally. Lowering the entry cost does not change this constraint. High productivity entrepreneurs who are still in the early stages of learning also opt to stay informal, preferring to wait until they have more precise beliefs about their profitability before formalizing. As a result, the entry cost reduction primarily influences a narrow group of entrepreneurs who already hold relatively accurate beliefs about their business potential and are close to formalizing. Consequently, the overall reduction in informality remains limited.

The effect of reducing entry costs is further limited in general equilibrium. Lower entry costs raise the expected value of entrepreneurship, encouraging more individuals to enter the entrepreneurial sector. In partial equilibrium, this results in a nearly 3% increase in the number of entrepreneurs. However, as more agents switch to entrepreneurship, the labor supply contracts. In general equilibrium, this labor scarcity puts upward pressure on wages, which rise from 1.66 to 1.67, a 0.31% increase. The higher wage raises the opportunity cost of entrepreneurship and increases the value of remaining a worker. As a result, the increase in entrepreneurial activity is smaller in general equilibrium: the number of entrepreneurs rises by only 0.08%. Consequently, The reduction in informality is more muted than in partial equilibrium, and the resulting gains in entrepreneurial output and tax revenue are modest, only 0.22% and 0.69%, respectively. These modest gains are not enough to offset the production loss in the corporate sector due to reduced capital and labor. As a result, aggregate output in the overall economy declines slightly, by 0.04%.

	Baseline	Reduced C^f_{entry} PE	GE
Prices			
r	3.99%		3.95%
w	1.66		1.67
Informality			
Informal ents (mass)	100	97.97	95.48
Informal ents (share)	73.64%	70.13%	70.26%
Informal output (share)	20.41%	17.96%	18.69%
Aggregate Economy			
Num. of ents (mass)	100	102.86	100.08
K_e	100	106.13	101.05
K	100	104.32	99.62
Y_e	100	105.82	100.22
Y	100	102.99	99.96
Measured ent TFP	27.74	27.96	27.72
Tax revenues	100	105.18	100.69

Table 9. Effects of Reducing the Formal Sector Entry Cost

Notes: This table reports the results of a policy experiment that reduces the entry cost into the formal sector by 50%. (i) K_e denotes the capital used in the entrepreneurial sector, while K refers to aggregate capital in the entire economy, including both the entrepreneurial and corporate sectors. (ii) Y_e and Y represent output produced by entrepreneurs and total output in the economy, respectively. (iii) Measured entrepreneurial TFP is calculated as $\frac{Y_e}{(K_e^{\alpha} N_e^{1-\alpha})^{\eta}}$, where Y_e, K_e, N_e are the aggregate entrepreneurial output, capital, labor, respectively.

6.2. Reducing Entry Cost and a Tax Relief

A growing body of evidence suggests that formalization policies are more effective when they reduce the ongoing costs of operating formally or increase the benefits associated with formality. Using Brazilian data, Rocha et al. (2018) provide clear evidence in this direction. They study a large scale formalization program implemented in two stages. The first stage focuses exclusively on reducing entry costs into the formal sector and generates only modest reductions in informality. In contrast, the second stage combines entry cost reductions with a tax relief for small formal firms and leads to substantially larger declines in informal activity. Importantly, the expansion of the formal sector is driven primarily by the formalization of existing informal firms rather than by the creation of new formal businesses.

Broadly following the structure of this reform, we consider a policy counterfactual that combines a 50% reduction in the fixed cost of formal sector entry with temporary tax relief for young formal firms. Specifically, formal firms aged 0 to 3 are exempt from paying the social security component of the payroll tax, reducing the payroll tax rate for these firms from 0.29 to 0.09.

	Baseline	Reduced C^f_{entry} + Tax cut PE	GE
Prices			
r	3.99%		3.80%
w	1.66		1.69
Informality			
Informal ents (mass)	100	91.38	79.55
Informal ents (share)	73.64%	59.41%	60.04%
Informal output (share)	20.41%	11.24%	11.40%
Aggregate Economy			
K_e	100	125.41	108.83
K	100	120.18	102.68
Y_e	100	125.61	107.22
Y	100	113.30	103.56
Measured ent TFP	27.74	28.95	28.04
Tax revenues	100	121.94	108.53
Occupation			
Num. of ents (mass)	100	113.27	97.56
Num. of q_{max} ents (mass)	100	111.16	95.25
Frac. of q_{max} ents who are formal	54.87%	59.18%	59.60%
Num. of formal q_{max} ents (mass)	100	119.91	103.46
Transition			
Num. of entrants to informal sec. (mass)	100	115.30	100.10
Num. of entrants to formal sec. (mass)	100	100	100
Num. of inf $\rightarrow$ f (mass)	100	144.29	122.31
Share of inf ent who transition	6.87%	10.83%	10.55%
Years to transition	3.62	2.14	2.14

Table 10. Effects of Reducing the Formal Sector Entry Cost and a Tax Relief

Notes: This table presents the results of the policy experiment, which reduces the entry cost to the formal sector by 50% and eliminate the contribution to social security for all firms in the formal sector with age ≤ 3. (i) K_e denotes the capital used in the entrepreneurial sector, while K refers to aggregate capital in the entire economy, including both the entrepreneurial and corporate sectors. (ii) Y_e and Y represent output produced by entrepreneurs and total output in the economy, respectively. (iii) Measured entrepreneurial TFP is calculated as $\frac{Y_e}{(K_e^\alpha N_e^{1-\alpha})^\eta}$, where Y_e, K_e, N_e are the aggregate entrepreneurial output, capital, labor, respectively.

Table 10 presents the results of this policy experiment. In the partial equilibrium setting, where prices are held fixed, the reform leads to a substantial reduction in informality. The share of informal entrepreneurs declines from 74% to 59%. The total number of informal entrepreneurs falls by nearly 9%, and the informal sector's share of entrepreneurial output drops by 9 percentage points, from 20% in the benchmark economy to 11% in the new steady state.

Beyond reducing informality, the policy also delivers sizable gains in aggregate economic performance. Total output increases by more than 13%, and measured entrepreneurial TFP rises by 4%, from 27.74 to 28.95. Importantly, despite including a tax cut, the policy is self-financing: overall tax revenue increases by nearly 22%.

The mechanism behind these gains is a reduction in occupational misallocation. By lowering the entry barrier and offering temporary tax relief to young formal firms, the policy raises the expected return to entrepreneurship, encouraging more individuals to enter the entrepreneurial sector. In the new steady state, the number of entrepreneurs increases by 13%. This is particularly valuable in an environment with information frictions, where individuals must experiment with entrepreneurship to learn their business quality. A higher entrepreneurship rate improves the discovery of high-productivity entrepreneurs. As a result, the number of entrepreneurs with the highest business quality realization (q_{max}) is 11% higher in the new steady state, and the share of these high-q entrepreneurs operating formally rises from 55% in the benchmark to nearly 60%.

Notably, the increase in entrepreneurship is driven by higher entry into informal entrepreneurship. The number of new informal entrepreneurs rises by 15%, while direct entry into the formal sector remains unchanged. This is because, despite the payroll tax cut, startups in the formal sector are still subject to a sales tax. Relative to the informal sector, formal entry remains costly. As a result, individuals continue to use the informal sector as a platform to test business ideas and form preliminary beliefs about their business quality before deciding whether to formalize. The decline in informality is therefore primarily driven by the formalization of existing informal firms, rather than the creation of new formal businesses, a pattern consistent with the empirical findings of Rocha et al. (2018).

To provide further evidence, we examine transitions from informal to formal entrepreneurship. In the new steady state, the number of entrepreneurs making this transition increases by 44%. The share of informal entrepreneurs who formalize also rises, from about 7% in the benchmark to nearly 11%. This acceleration in formalization is driven by a reduction in the cost of learning within the formal sector. With lower entry costs and temporary tax relief, entrepreneurs no longer need highly accurate beliefs about their business quality before formalizing. Instead, they shorten the experimentation phase in the informal sector and shift part of the learning process into the formal sector. In the benchmark economy, it takes an average of 3.62 years for an entrepreneur to formalize; under the reform, this duration falls to just 2.14 years. Many young firms choose to formalize earlier to take advantage of the tax exemption. Under the policy, entrepreneurs begin informally, form a preliminary signal about their business potential, and then complete the remainder of the learning process in the formal sector, a shift made possible by the temporary reduction in tax obligations.

Turning to the general equilibrium outcomes, the wage increases by 1.42%, from 1.66 to 1.69. The higher wage reduces individuals' incentive to become entrepreneurs. As a result, the number of new informal sector entrants rises only marginally, by 0.1%. It also causes more young firms to exit by raising the value of wage employment. Entrepreneurs in the early stages of learning who experience a negative transitory shock are now more likely to exit and take up wage work. Consequently, the total number of entrepreneurs falls in general equilibrium, and there are fewer entrepreneurs with the highest business quality, $q_{\max}$. This means that the extensive margin, i.e., the output gains from increased occupational choice, is weakened by general equilibrium effects.

However, the intensive margin, i.e., the output gains from improved sectoral allocation of entrepreneurs, remains active. Lower entry costs and temporary tax relief reduce the cost of learning within the formal sector, encouraging entrepreneurs to shorten their learning period in the informal sector and formalize earlier. The average time to transition drops from 3.62 to 2.14 years. As a result, the number of informal entrepreneurs transitioning to the formal sector increases by 22%, leading to a 48% increase in the number of formal sector entrepreneurs. Informality declines substantially: the number of informal entrepreneurs falls by 20%, the share of informal entrepreneurs drops by 14 percentage points, and the informal sector's share of entrepreneurial output shrinks from 20% to 11%.

Even though the decline in total entrepreneurship leads to fewer entrepreneurs with the highest business quality ($q_{\max}$), the accelerated transition process brings more of them into the formal sector. In the new steady state, the number of formal sector entrepreneurs with $q_{\max}$ increases by nearly 3.5% compared to the baseline. These high-productivity entrepreneurs benefit from operating formally, where they enjoy better access to capital and face no labor distortions, allowing them to scale up production more efficiently. As a result, total output produced by high-productivity entrepreneurs rises by 18%. In addition, measured entrepreneurial productivity increases by 1%, overall entrepreneurial output grows by more than 7%, and total output expands by nearly 4%. Tax revenue also rises by over 8%, rendering the policy self-financing. Notably, the tax relief is temporary, limited to the first four years of a firm's life, yet it still generates sizable gains for the aggregate economy. This highlights a key policy implication: in developing countries, targeted support for young firms can be an effective strategy for boosting output and reducing informality.

6.3. Increasing the Monitoring Effort

Lastly, we consider a policy that raises the cost of operating informally, which can be interpreted as a result of increased government monitoring. In the model, this is implemented by reducing the parameter b, which governs the labor distortion in the informal sector. More precisely, we cut b in half by multiplying it by 0.5.

Table 11 presents the results of increasing the cost for informal entrepreneurs. Starting with the partial equilibrium case, the policy significantly reduces informality: the number of informal entrepreneurs falls by nearly 50%, and their share among all entrepreneurs declines from 74% to 53%. Similarly, the informal sector's share of entrepreneurial output drops sharply, from 20% to 5%.

	Baseline	Stronger Monitoring PE	GE
Prices			
r	3.99%		4.20%
w	1.66		1.64
Informality			
Informal ents (mass)	100	53.39	60.63
Informal ents (share)	73.64%	53.42%	53.16%
Informal output (share)	20.41%	5.61%	5.26%
Aggregate Economy			
K_e	100	83.08	93.57
K	100	86.03	96.78
Y_e	100	85.30	97.10
Y	100	92.23	98.95
Measured ent TFP	27.74	27.27	27.78
Tax revenues	100	95.88	104.87
Occupation			
Num. of ents (mass)	100	73.60	83.98
Num. of q_{max} ents (mass)	100	85.51	89.79
Frac. of q_{max} ents who are formal	54.87%	64.73%	68.52%
Num. of formal q_{max} ents (mass)	100	100.87	112.12
Transition			
Num. of entrants to informal sec. (mass)	100	79.77	90.10
Num. of entrants to formal sec. (mass)	100	100	100
Num. of inf $\rightarrow$ f (mass)	100	102.12	118.00
Share of inf ent who transition	6.87%	13.09%	13.32%
Years to transition	3.62	1.96	2.00

Table 11. Effects of Enhancing Government Monitoring of the Informal Sector

Notes: This table presents the results of the policy experiment, which increases government monitoring efforts in the informal sector. In our model, this is implemented by multiplying the parameter b by 0.5. (i) K_e denotes the capital used in the entrepreneurial sector, while K refers to aggregate capital in the entire economy, including both the entrepreneurial and corporate sectors. (ii) Y_e and Y represent output produced by entrepreneurs and total output in the economy, respectively. (iii) Measured entrepreneurial TFP is calculated as $\frac{Y_e}{(K_e^{\alpha} N_e^{1-\alpha})^{\eta}}$, where Y_e, K_e, N_e are the aggregate entrepreneurial output, capital, labor, respectively.

Despite its success in reducing informality, the policy imposes substantial economic costs in our model. Aggregate output declines by nearly 8%, measured entrepreneurial TFP falls by 1.7%, and total tax revenue drops by more than 4%.[13] These results stand in contrast to those in Ulyssea (2020), where increased government enforcement not only reduces informality but also boosts aggregate output. The divergence stems from the role played by the informal sector. In our framework, the informal sector functions as a stepping stone: due to information frictions, it offers entrepreneurs a low-cost environment to experiment and learn about their business quality. Disrupting this platform restricts entrepreneurial experimentation and discourages entry into entrepreneurship, ultimately harming economic performance.

In general equilibrium, the decline in entrepreneurship increases the supply of labor. As a result, the wage rate falls by 1.58%, from 1.66 to 1.64. The lower wage mitigates the negative effect on entrepreneurial entry. Compared to the partial equilibrium results, the total number of entrepreneurs and new informal sector entrants decline less. In the new steady state, the number of entrepreneurs decreases by 16%, and the number of new informal entrepreneurs falls by about 10%. With fewer entrepreneurs, aggregate capital declines, leading to a 3% drop in entrepreneurial output and a 1% reduction in total output.

As in the previous policy experiment, entrepreneurs in the informal sector speed up their transition to formality. However, the mechanism here is different. Entrepreneurs do not formalize early because learning becomes less costly in the formal sector. Instead, they do so because the cost of remaining informal becomes too high. Once they develop a rough belief that their business quality is likely to be high, they choose to formalize quickly to avoid these rising costs.

This policy generates two opposing effects. On the one hand, it reduces the space for entrepreneurial experimentation, thereby discouraging new business formation and harming aggregate economic performance. On the other hand, it pushes more informal entrepreneurs to formalize sooner, improving resource allocation. For instance, the number of formal sector entrepreneurs with the highest business quality increases by 12% in the new steady state. Ultimately, the negative effects dominate. The intuition is straightforward: in the presence of learning frictions, the informal sector acts as a pipeline for future formal entrepreneurs. Eliminating this low-cost experimentation platform undermines entrepreneurial discovery and weakens long-term economic performance.

7. Conclusion

This paper examines the role of the informal sector as a stepping stone for entrepreneurship by explicitly incorporating business learning and uncertainty into a model of occupational and sectoral choice. In our framework, entrepreneurs are uncertain about their underlying business quality and must learn through production outcomes. The informal sector provides a low-cost environment for this experimentation, allowing entrepreneurs with high productivity business idea to emerge and, over time, transition into the formal sector.

[13]Note that the actual fiscal cost could be higher, as we abstract from any administrative or enforcement costs associated with increased monitoring. In reality, implementing such policies is likely to be costly, especially given the difficulty of monitoring a large number of small firms.

Our model highlights the interaction between entrepreneurial learning, occupational decisions, and sectoral transitions. We use this framework to evaluate the macroeconomic effects of various policies. We begin by analyzing a policy that reduces the formal sector entry cost, and then consider a more comprehensive reform that combines entry cost reductions with temporary tax relief for young formal sector firms. The model replicates empirical patterns documented in recent studies: cutting entry costs alone has limited effects on informality, while reducing the ongoing costs of operating formally leads to much greater formalization.

We also study the effects of intensifying government monitoring of the informal sector. While this policy substantially reduces informality, both in terms of the number of informal entrepreneurs and the informal sector's share of output, it also generates important trade-offs. In our model with learning, restricting access to informal experimentation discourages business entry, reduces total entrepreneurship, and increases occupational misallocation. These distortions lead to unintended macroeconomic consequences.

This paper contributes to the literature on informality and entrepreneurship by offering a new perspective on the informal sector, not merely as a distortion, but as a mechanism for identifying high-productivity entrepreneurs who might otherwise remain undiscovered. A key policy implication is that efforts to restrict informal activity may inadvertently suppress the emergence of productive firms, imposing long-run economic costs. Our findings suggest that effective formalization policies must strike a careful balance: rather than focusing solely on eliminating informality, policies should support young firms by providing an effective platform for business learning.

This paper relies on two key simplifying assumptions. First, we do not allow firms that have entered the formal sector to transition back into informality. Second, we assume homogeneous priors across all individuals. While both assumptions could be relaxed in future work and represent natural extensions of the framework, they do not undermine our main results or the core mechanisms emphasized in this paper. Our main contribution is to demonstrate the positive role the informal sector can play in the presence of information frictions, particularly for entrepreneurs who turn out to have high business productivity. Given this focus, these simplifications help highlight the learning channel and clarify how informality facilitates entrepreneurial experimentation.

We acknowledge that, in the data, some formal firms do transition to informality, with the likelihood of such transitions declining with the entrepreneur's income. This pattern could be captured in an extended version of the model that introduces heterogeneity in prior beliefs and allows formal firms to switch to the informal sector. In such a setting, overconfident entrepreneurs may choose to start their businesses formally, only to shift to the informal sector after discovering that their productivity is lower than expected. We abstract from this channel in order to maintain a sharp focus on our main research question: how information frictions shape the role of informality in entrepreneurial dynamics. Even with this additional margin, the main conclusions from our policy counterfactuals would continue to hold. Lowering the cost of entry into the formal sector would still have limited effect, as entrepreneurs sort into sectors based on their beliefs. Similarly, a temporary payroll tax relief would still encourage transitions from informal to formal status, as many entrepreneurs shift part of their learning process to the formal sector once they gain an initial signal. As a result, the policy would still yield gains in aggregate output and formalization.

Moreover, allowing for transitions from formality back into informality would not alter the key result that increased enforcement generates negative effects on the economy. In addition to discouraging individuals from entering the informal sector to experiment and learn about their business potential, stricter monitoring would also reduce the fallback option for entrepreneurs who initially enter the formal sector with optimistic priors but later discover low productivity. If moving to informality becomes more costly or less feasible, the value of experimentation diminishes, leading more individuals to forgo entrepreneurship altogether and remain in wage employment. As a result, the overall negative effect on the economy would likely persist.

One further assumption worth discussing is our adoption of a simple Bayesian learning structure with homogeneous belief updating across all entrepreneurs, regardless of sector. This abstraction allows us to isolate and emphasize the core mechanism at play. That said, an open and compelling question is whether entrepreneurs in different sectors adopt different learning strategies. For example, do those operating in the formal and informal sectors differ in how they acquire information and how quickly they learn? A deeper understanding of these dynamics would help policymakers design more targeted interventions that support productive experimentation without prematurely crowding it out. While modeling such heterogeneity lies beyond the scope of this paper, it represents an important and promising direction for future research.

References

Allub, L. and Erosa, A. (2019). Financial frictions, occupational choice and economic inequality. *Journal of Monetary Economics*, 107:63–76.

Bruhn, M. and McKenzie, D. (2014). Entry regulation and the formalization of microenterprises in developing countries. *The World Bank Research Observer*, 29(2):186–201.

Buera, F. J., Kaboski, J. P., and Shin, Y. (2011). Finance and development: A tale of two sectors. *American economic review*, 101(5):1964–2002.

Buera, F. J. and Shin, Y. (2013). Financial frictions and the persistence of history: A quantitative exploration. *Journal of Political Economy*, 121(2):221–272.

Chen, C., Senga, T., Sun, C., and Zhang, H. (2023). Uncertainty, imperfect information, and expectation formation over the firm's life cycle. *Journal of Monetary Economics*, 140:60–77.

Conesa, J. C., Kitao, S., and Krueger, D. (2009). Taxing capital? not a bad idea after all! *American Economic Review*, 99(1):25–48.

David, J. M., Hopenhayn, H. A., and Venkateswaran, V. (2016). Information, misallocation, and aggregate productivity. *The Quarterly Journal of Economics*, 131(2):943–1005.

De Mel, S., McKenzie, D., and Woodruff, C. (2008). Returns to capital in microenterprises: evidence from a field experiment. *The quarterly journal of Economics*, 123(4):1329–1372.

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

Djankov, S., La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2002). The regulation of entry. *The quarterly Journal of economics*, 117(1):1–37.

Erosa, A. (2001). Financial intermediation and occupational choice in development. *Review of Economic Dynamics*, 4(2):303–334.

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

Fajardo, J., Ornelas, J. R. H., and Farias, A. R. d. (2012). Estimating risk aversion, risk-neutral and real-world densities using brazilian real currency options. *Economia Aplicada*, 16:567–577.

Feng, Y. (2025). Firm life-cycle learning and misallocation. *Working Paper*.

Feng, Y., Lagakos, D., and Rauch, J. E. (2024). Unemployment and development. *The Economic Journal*, 134(658):614–647.

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

Gao, H. and Zhang, L. (2024). Uncover "gazelles": The macroeconomic implications of uncertainty and learning for entrepreneurship. *Working Paper*.

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

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

Guvenen, F., Kambourov, G., Kuruscu, B., Ocampo, S., and Chen, D. (2023). Use it or lose it: Efficiency and redistributional effects of wealth taxation. *The Quarterly Journal of Economics*, 138(2):835–894.

Herreño, J. and Ocampo, S. (2023). The macroeconomic consequences of subsistence self-employment. *Journal of Monetary Economics*, 136:91–106.

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

Jaar, D. (2026). Self-employment as self-insurance. *Available at SSRN 5175208*.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica: Journal of the econometric society*, pages 649–670.

Jung, J. and Tran, C. (2012). The extension of social security coverage in developing countries. *Journal of Development Economics*, 99(2):439–458.

Kaplan, D. S., Piedra, E., and Seira, E. (2011). Entry regulation and business start-ups: Evidence from mexico. *Journal of Public Economics*, 95(11-12):1501–1515.

Karlan, D. and Zinman, J. (2010). Expanding credit access: Using randomized supply decisions to estimate the impacts. *The Review of Financial Studies*, 23(1):433–464.

La Porta, R. and Shleifer, A. (2014). Informality and development. *Journal of economic perspectives*, 28(3):109–126.

Maloney, W. F. (2004). Informality revisited. *World development*, 32(7):1159–1178.

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

Rocha, R., Ulyssea, G., and Rachter, L. (2018). Do lower taxes reduce informality? evidence from brazil. *Journal of development economics*, 134:28–49.

Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics letters*, 20(2):177–181.

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

Ulyssea, G. (2020). Informality: Causes and consequences for development. *Annual Review of Economics*, 12(1):525–546.

Appendices

A. PNAD

The PNAD (Pesquisa Nacional por Amostra de Domicílios) is a nationally representative annual household survey conducted by IBGE. The data are publicly available on the IBGE website.[14] We use PNAD data from 2002 to 2009, pooling all years and restricting the sample to non agricultural sector individuals aged 18 to 65. We exclude individuals who are either unemployed or not in the labor force.

In our final sample, 57% of individuals are male and 43% are female. Approximately 95.13% of respondents report having only one job, and an additional 4.44% report having two jobs. Fewer than 0.5% report holding three or more jobs. We focus exclusively on individuals' main job. PNAD explicitly distinguishes between the main job, second job, and any other jobs, so no imputation is needed to identify the primary occupation.

Figure 10 displays the annual share of entrepreneurs within the non-agricultural working population, along with the share of informal entrepreneurs among all entrepreneurs. Both statistics remain relatively stable across survey years. In the pooled sample, 25% of individuals are classified as entrepreneurs, and among them, 26% operate in the formal sector.

For income, we use two key variables: (i) monthly effective income from the main job, and (ii) total monthly effective income from all sources. For entrepreneurs, we treat the effective income from their main job as their monthly business profit. The income distribution used throughout the analysis is based on total income from all sources. On average, 94% of individuals' total income comes from their main job. Even among those in the top income decile, income from the main job accounts for more than 87% of total income.

The average age of individuals in the full sample is 38 years. Disaggregating by occupational status, the average age of workers is 36, while that of entrepreneurs is 41. Among entrepreneurs, those in the informal sector are slightly younger on average (40 years) than their formal sector counterparts (43 years). Entrepreneurs in the top income decile are older, with both informal and formal entrepreneurs averaging 44 years of age. Figure 11 presents the average age of informal and formal entrepreneurs across income deciles.

Turning to firm age, the mean across all firms is 9 years. Informal sector businesses are, on average, three years younger than formal sector firms, with average ages of 8 and 11 years, respectively. Among firms owned by entrepreneurs in the top income decile, the average age of informal firms is 10 years, while formal firms average 12 years. Figure 12 displays the average firm age for informal and formal entrepreneurs by income decile.

[14] `https://www.ibge.gov.br/en/statistics/social/labor/20620-summary-of-indicators-pnad2.html`

As pointed out by prior studies (e.g., Ulyssea (2020), Maloney (2004)), formal and informal firms coexist within industries, and the two sectors are deeply integrated rather than segmented. We further confirm this pattern using PNAD data. In Figure 13, the upper panel shows the share of entrepreneurs across industries, while the lower panel depicts the informality composition within each industry. In every industry, both formal and informal entrepreneurs are present. There is no sector exclusively dominated by one type. On average, formal sector entrepreneurs represent approximately 26% of all entrepreneurs, and this proportion remains relatively consistent across most industries.

Similar patterns hold when we restrict the sample to entrepreneurs in the top income decile. Figure 14 focuses on this group. The upper panel displays the industry distribution of top-decile entrepreneurs. Compared to the full sample, their industry composition differs somewhat. While the three most common industries for all entrepreneurs are trade and retail, construction, and manufacturing, high-income entrepreneurs are much less likely to work in construction. Instead, 22% are concentrated in a category labeled "other industrial activities." Despite these differences in sectoral composition, we continue to observe a high degree of integration between informal and formal entrepreneurs within industries. The lower panel shows the composition of entrepreneurs by formality within each sector. On average, 63% of top-decile entrepreneurs are formal, and in most industries, formal entrepreneurs account for approximately 60% of the group—indicating no substantial within-industry differences in formality status among high-income entrepreneurs.

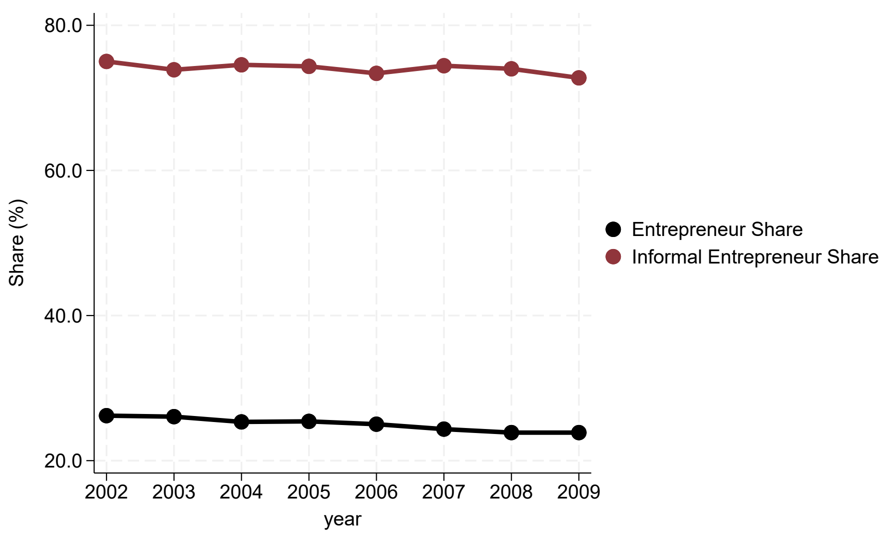

Figure 10. Entrepreneur Share and Informal Sector Entrepreneur Share in Brazil: 2002-2009

Notes: Data from PNAD. This figure displays two time series from 2002 to 2009: the share of entrepreneurs in the non-agricultural working population, and the share of informal entrepreneurs among all entrepreneurs.

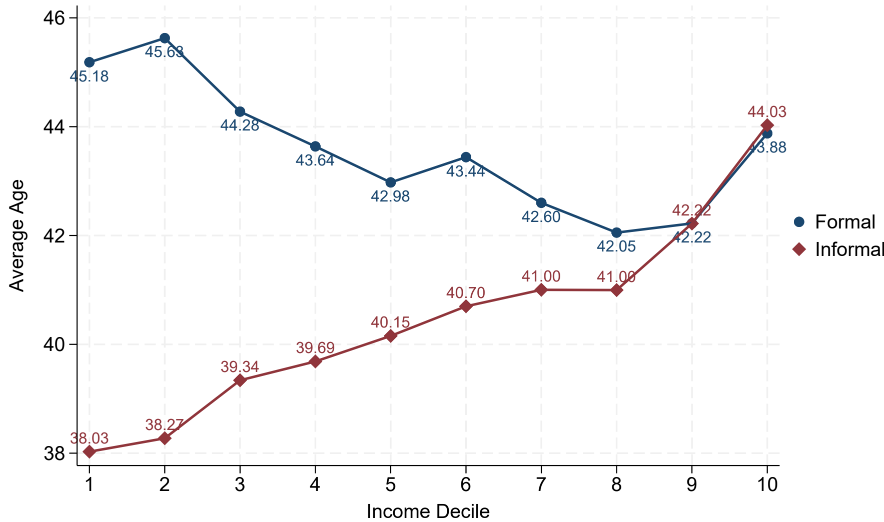

Figure 11. Average Age of Entrepreneurs in Each Income Decile

Notes: Data from PNAD. This figure shows the average age of informal entrepreneurs (red) and formal entrepreneurs (blue) in each income decile.

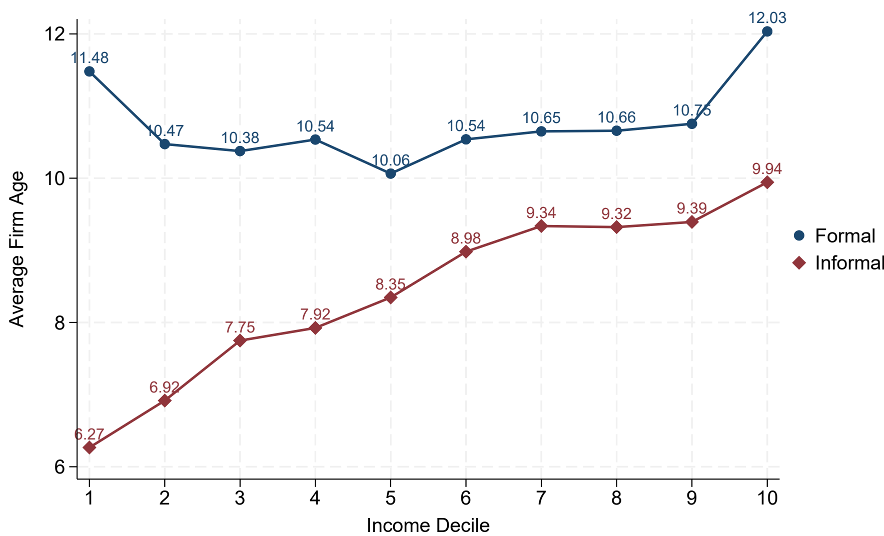

Figure 12. Average Age of Firms Operated by Entrepreneurs Across Income Deciles

Notes: Data from PNAD. This figure shows the average age of firms operated by informal (red) and formal entrepreneurs (blue) in each income decile.

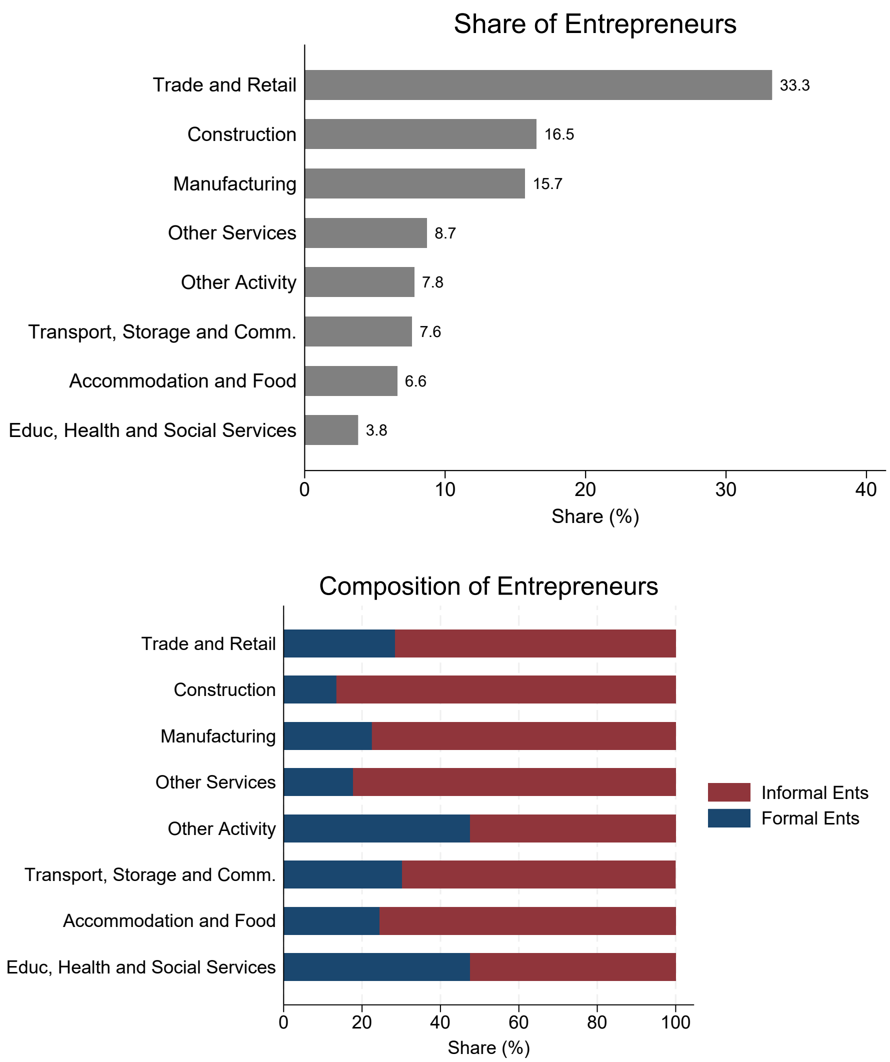

Figure 13. All Entrepreneurs: Industry Shares and Informality Composition

Notes: Data from PNAD. The upper panel shows the share of entrepreneurs in each industry. The lower panel displays, within each industry, the composition of entrepreneurs: specifically, the share of informal entrepreneurs (red) and the share of formal entrepreneurs (blue).

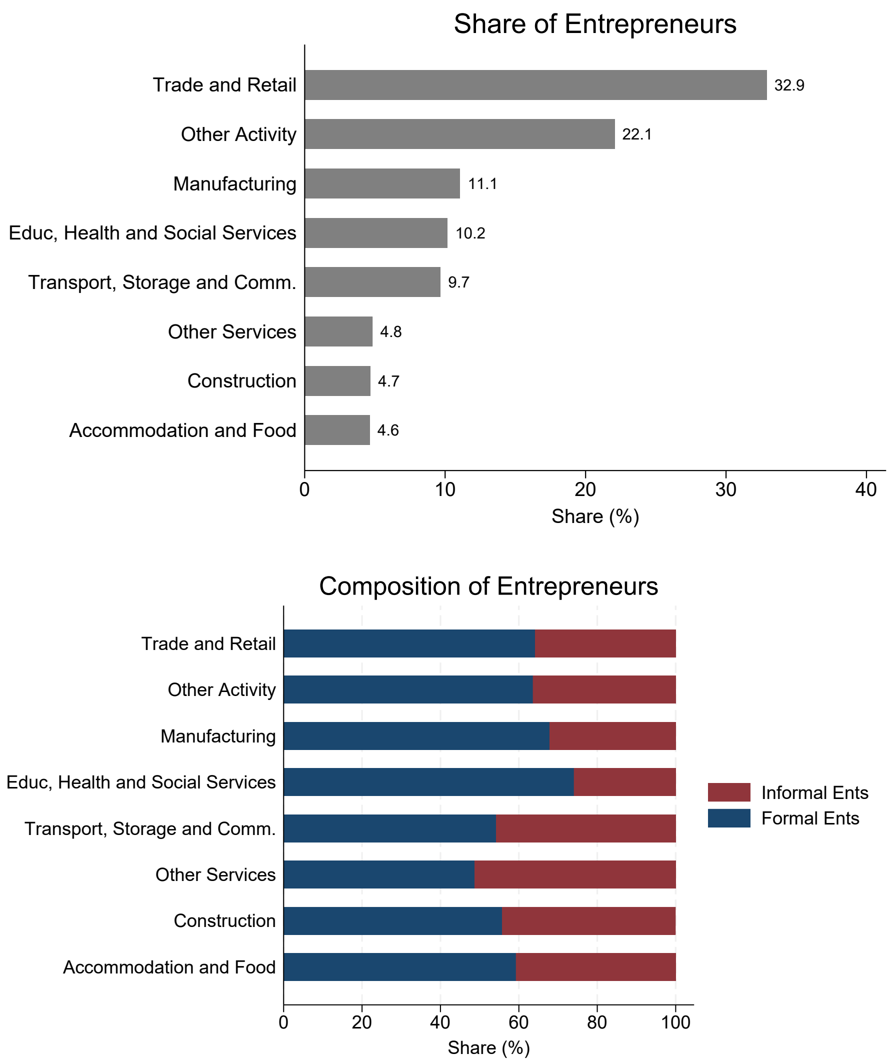

Figure 14. Top Income Decile Entrepreneurs: Industry Shares and Informality Composition

Notes: Data from PNAD. This figure figure focuses on top income decile entrepreneurs only. The upper panel shows the share of entrepreneurs in each industry. The lower panel displays, within each industry, the composition of entrepreneurs: specifically, the share of informal entrepreneurs (red) and the share of formal entrepreneurs (blue).

B. ECINF

The ECINF survey (Pesquisa de Economia Informal Urbana) is a repeated cross-sectional survey of small firms, conducted by IBGE in 1997 and 2003. It is publicly available on IBGE's website.[15] ECINF is a matched employer-employee dataset that provides detailed information on entrepreneurs, their businesses, and their employees. It is designed to be nationally representative of small urban firms, covering both formal and informal businesses. In our analysis, we use the most recent wave of the survey, collected in 2003.

One potential concern with ECINF is that it may not fully capture the characteristics of formal sector firms. Since the survey is limited to small firms with a cap on the number of employees, it is well suited for studying the informal sector, where most firms are indeed small, but may underrepresent the formal sector by excluding larger firms. However, Ulyssea (2020) compares ECINF with RAIS, a high quality administrative dataset that covers the universe of formal firms in Brazil. The study finds that both the size distribution and the industry composition of formal firms in ECINF closely resemble those in RAIS, which is reassuring regarding the reliability and representativeness of ECINF data for analyzing formal firms.

In the PNAD data, 26% of entrepreneurs operate in the formal sector, compared to 24% in ECINF. Among entrepreneurs in the top income decile, 37% are in the informal sector in PNAD, while the corresponding figure in ECINF is 42%. Table 12 provides a comparison of key statistics across the two datasets, including entrepreneur characteristics and firm attributes. Overall, the patterns are highly consistent, indicating that PNAD and ECINF offer broadly comparable insights into the structure of entrepreneurship in Brazil.

Variable	PNAD	ECINF
Share of ents who are in the top income decile	15	16
Ent age (median / mean)	41 / 41	41 / 41
Informal ent age (median / mean)	40 / 40	40 / 41
Formal ent age (median / mean)	43 / 43	42 / 42
Share of informal ent with a college degree or above	5	6
Share of top income decile informal ent with a college degree or above	31	32
Firm age (median / mean)	6 / 9	6 / 9
Informal firm age (median / mean)	5 / 8	6 / 9
Formal firm age (median / mean)	9 / 11	7 / 9

Table 12. PNAD vs ECINF

Notes: This table reports the summary statistics from PNAD and ECINF.

[15] https://www.ibge.gov.br/en/statistics/social/labor/16845-urban-informal-economy.html?edicao=16847

C. PNADC

To compute transitions from informal to formal entrepreneurship, we use data from the PNAD Contínua (PNADC), which is publicly available on the IBGE website.[16] Launched by IBGE in 2012, PNADC is a rotating panel survey that tracks households over five consecutive quarters. It shares a similar structure with the original PNAD survey but introduces a panel component that allows for short-run dynamic analysis. In 2015, IBGE officially phased out PNAD and fully replaced it with PNADC.

The 2012–2013 waves of PNADC are usually regarded as a transitional series due to the survey's gradual rollout and ongoing methodological adjustments during that period. For instance, a technical report by researchers at Ipea noted that "[PNAD Contínua's] data collection was still under development at that time."[17] By 2015, PNADC had achieved methodological and sample design stability and became Brazil's sole official household survey. In addition, Brazil experienced an economic recession from mid-2014 to mid-2016, with the economy officially recovering in early 2017. To avoid periods affected by data reliability concerns or macroeconomic volatility, we base our transition analysis on the 2017 wave of PNADC.

Most individuals in our sample report having only one job: nearly 97% report a single job, while an additional 3% report having two jobs. Only about 0.25% report holding three or more jobs. We use information on individuals' main job. In PNADC, the share of entrepreneurs within the non-agricultural working population is 28%, which is close to the 25% found in PNAD. The share of informal sector entrepreneurs in PNADC is also similar: 28% compared to 26% in PNAD. Moreover, we find that around 30% of top income decile entrepreneurs in PNADC operate in the informal sector, consistent with our main finding that one-third of high-income entrepreneurs run informal businesses. Table 13 presents the firm size distribution from both PNAD and PNADC, which appear quite similar.

Firm Size	PNAD			PNADC		
	All Ents	Informal	Formal	All Ents	Informal	Formal
0	81.32	90.49	55.18	84.48	95.70	56.02
1-5	13.83	8.17	29.99	11.26	3.92	29.91
6-10	2.40	0.77	7.03	2.16	0.25	7.00
≥ 11	2.45	0.57	7.80	2.09	0.13	7.07

Table 13. Firm Size Distribution: PNAD vs PNADC

Notes: This table reports the distribution of firm size for all entrepreneurs, as well as separately for informal and formal sector entrepreneurs, using both PNAD and PNADC data.

[16] https://www.ibge.gov.br/en/statistics/social/labor/2866-np-continuous-national-househol 31241-pnadc-pnadc4-en.html

[17] repositorio.ipea.gov.br

As PNADC is a household-level survey, it does not track firms directly. Instead, we observe the formality status of the business operated by each individual based on their reported CNPJ registration status. However, we cannot confirm whether it is the same business across survey quarters. As a result, a potential concern in analyzing transitions between informal and formal entrepreneurship is that some observed transitions may reflect the closure of one business and the opening of a new one, rather than the formalization of an existing enterprise.

To address this issue, we implement the following procedure. First, we examine the number of sectoral transitions per individual. We find that frequent transitions are relatively rare: 90% of the sample either experience no transition across sectors (worker, informal entrepreneur, formal entrepreneur) or only transition once. About 8% of individuals have two transitions, and only around 2% have three or more. To ensure clean identification of transitions, we restrict the sample to individuals who have either no transition or only one. Among those who begin the panel as entrepreneurs, we track their firm's reported age and industry across quarters. This helps us verify whether the individual continues to operate the same business, such that any observed change in formality status can be reasonably interpreted as a transition of the same firm.

The average age of firms at the time of formalization is seven years, indicating that, on average, it takes informal entrepreneurs seven years to transition to the formal sector. To investigate the factors associated with the speed of this transition, we restrict the sample to entrepreneurs who move from informal to formal status and regress the number of years until formalization on a range of individual and firm characteristics. Table 14 presents the estimation results. We find no statistically significant relationship between transition duration and individual characteristics, most notably, years of schooling, a standard proxy for individual ability.

Years	
Education	-0.034
Male	1.074*
Age	0.202
Age squared	0.003*
Firm size	-0.002
Race	Yes
State	Yes
Industry	Yes
Observations	1,525
R-squared	0.404

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 14. Years to Transition from Informal to Formal Entrepreneurship

Notes: Data from PNADC. The table reports regression results on the number of years it takes for entrepreneurs to transition from informal to formal status, controlling for both individual characteristics and firm-level attributes.