Entrepreneurship and Informality in Brazil

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

[1]York University

[2]University of North Carolina at Chaptel Hill

Abstract. *Most developing countries have a substantial informal sector, yet its role in entrepreneurship and macroeconomic outcomes remains underexplored. Understanding the causes and consequences of informality is crucial for economic growth, welfare, and policy design. This paper examines the relationship between informality and entrepreneurship, making three key contributions to the literature. First, we document new empirical facts about informality and entrepreneurship, showing that (i) a significant share of high-income entrepreneurs operate their business in the informal sector, (ii) formal sector entrepreneurs are concentrated in higher income groups, and (iii) high-income informal sector entrepreneurs are more likely to transition to the formal sector over time. Second, motivated by these empirical findings, we develop a theoretical model that incorporates imperfect information and learning, allowing individuals to choose between wage work and entrepreneurship without fully knowing their business potential. This framework highlights how informality serves as a platform for entrepreneurial experimentation. Third, we use the calibrated model to evaluate the impact of formalization policies, providing insights relevant to both academic research and policy debates.*

Keywords*: Informality, Entrepreneurship, Learning, Occupation Misallocation, Growth, Formalization Policy*
JEL Codes*: E26, H20, L26, O17, D80*

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. Importantly, these businesses and individuals do not engage in illicit activities, such as drug trafficking or human smuggling; rather, they operate without formal business registration or labor contracts. As a result, a substantial fraction of economic activity in developing countries is completely unregulated, taking place at the margin of tax, labor and other relevant regulatory frameworks. This extensive informal sector has profound

*Email: yrguo@yorku.ca

We are grateful to Tasso Adamopoulos, Chaoran Chen, Lutz Hendricks, Gueorgui Kambourov, Eugene Tan, Can Tian, and all participants at UNC, UofT, York, and SEA 2024 Annual Meeting for their questions, comments, and insights.

implications for individual behavior, which reverberate throughout the economy, impacting productivity, aggregate output, and economic growth. Thus, understanding the causes and consequences of informality is vital for both academic research and effective policymaking.

Using data from Brazil's National Household Sample Survey (PNAD), which provides detailed information on households in both the formal and informal sectors, we document three empirical facts that are new to the literature: (1) Over one-third of entrepreneurs in the top income decile operate businesses in the informal sector, challenging the conventional view that those in the informal sector lack the education or skills for formal employment, or are too poor to access public and financial services. (2) Entrepreneurs in the formal sector are predominantly concentrated in higher income groups, suggesting they may bypass the typical entrepreneurial path of starting small and gradually growing into the top income brackets. (3) High-income informal sector entrepreneurs are more likely to transition to the formal sector than their lower-income counterparts.

Motivated by these findings, we develop a theory of business learning and entrepreneurship that reconciles these empirical observations; ultimately, we seek to understand how the interaction between business learning and sectoral selection affects the overall economy of developing countries. The central idea is that uncertainty, arising from information friction, leads individuals to choose between wage work and entrepreneurship without fully knowing their entrepreneurial capabilities. Entrepreneurs, therefore, must utilize their production output in each period as a signal to update their beliefs about their abilities. In this context, the informal sector may function as a platform for entrepreneurship experimentation, as it allows for lower production costs by avoiding taxes and fixed operational expenses.

In summary, this research project aims to systematically integrate learning dynamics into the analysis of the informal sector's impact on the aggregate economy. Central to this investigation is the hypothesis that informality serves as a low-cost platform for households to experiment with entrepreneurial ideas, facilitating the discovery of high-potential entrepreneurs who might otherwise lack the resources and opportunities to test their capabilities. This experimentation is particularly crucial in developing economies, where access to capital and formal support structures can be severely constrained. However, while informality can foster entrepreneurial experimentation, it also imposes significant limitations that may restrict potential benefits. For instance, informal entrepreneurs often encounter stricter borrowing constraints, hindering their ability to scale operations or invest in new ventures. Additionally, they risk substantial penalties if identified by regulatory authorities. Thus, the intricate interplay between informality and learning creates a dynamic that can either enhance or impede efficient resource allocation and overall economic output. This complex relationship forms the basis of our quantitative research question: Does the interaction between informality and learning lead to improved or deteriorated economic outcomes?

To address this question, we, first, document the characteristics of entrepreneurs and their businesses across both informal and formal sectors using the Brazilian National Household Sample Survey (PNAD) data and the Urban Informal Economy Sur-

vey (ECINF) data. Then, building on the empirical observations, we develop a general equilibrium life-cycle model that incorporates heterogeneous agents. This model captures the critical occupation choices that households face: whether to engage in wage work or pursue entrepreneurship, and, if the latter, whether to operate within the formal or informal sector. While productivity associated with wage work is known, entrepreneurial ability is uncertain, requiring individuals to update their beliefs about their business potential through production outcomes. Production in the informal sector is associated with lower costs but also with tighter financial constraints and concerns over regulatory repercussions. In each period, entrepreneurs in the informal sector will have the option to transition to the formal sector, capturing the rich dynamics of entrepreneurial transitions observed from the data.

This theoretical model is then be estimated to match the key empirical moments documented in the empirical part. The calibrated framework serves as a laboratory for conducting policy experiments focused on the broader implications of informality for economic dynamics. Specifically, we analyze the effects of a policy that increases the cost of informal business operations, which is interpreted as intensified government monitoring. We focus on this formalization policy because prior research suggests that stricter enforcement is more effective at reducing informality than policies such as lowering formal sector entry costs. Previous literature finds that stricter monitoring can significantly reduce informality while boosting productivity and tax revenues (see, for example, Ulyssea (2018), Erosa et al. (2023)).

However, by explicitly modeling business learning, our paper has different results. We highlight the informal sector's role as a stepping stone for entrepreneurship, showing that raising the cost of informality can have unintended consequences. While the policy successfully reduces informality by cutting the number of informal sector entrepreneurs by 77% and reducing informal sector output by 9 percentage points, it also introduces significant distortions. Stricter enforcement raises barriers to business experimentation, leading to a 68% decline in total entrepreneurship and a 49.6% drop in formal sector entrepreneurs. Occupational misallocation worsens as highly productive individuals are discouraged from starting businesses, reducing the number of top-quality entrepreneurs by 26%. Unlike previous studies that emphasize the benefits of formalization policies, we show that restricting informal entry can undermine entrepreneurial selection and long-run business dynamism. These distortions generate large macroeconomic consequences: aggregate output falls by 33%, measured entrepreneurial productivity declines by 9%, and total tax revenue shrinks by 46%. The government must impose a lump-sum tax, further depressing household consumption and welfare, which declines by nearly 28%. Importantly, these negative effects arise precisely because business learning is curtailed, as potential entrepreneurs remain workers 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 directly builds on the literature in entrepreneurship and inequality (see, for example, Quadrini (2000), Cagetti and De Nardi (2006)), extending it to the context of developing economies, where the informal sector plays a defining role. By examining how choices between informal and formal sectors interact with learning, this paper provides new insights into how entrepreneurship contributes to inequality and income mobility in developing countries. Moreover, it addresses a foundational question concerning determinants of informal sector size—a topic explored both empirically and theoretically in prior studies. For instance, Busso et al. (2012) identify human capital as a key driver of informal sectors in developing economies, while Alonso-Ortiz and Leal (2018) and Levy (2008) examine the role of social programs and transfers in Mexico. Our model introduces the idea of informality as a learning mechanism, offering a new perspective on these established findings.

While this paper builds on existing studies that examine the macroeconomic implications of informality through the lens of entrepreneurship, it offers a distinct perspective that expands upon previous work. Notably, studies such as Franjo et al. (2022) and Erosa et al. (2023) have primarily concentrated on financial frictions as the main channel through which informality affects economic outcomes. In contrast, our paper integrates learning and experimentation as additional key factors influencing entrepreneurial decisions. By incorporating learning dynamics alongside financial frictions, our model provides a nuanced understanding of informality as a strategic choice for households navigating uncertainty in their entrepreneurial abilities.

Additionally, our framework highlights how participation in the informal sector affects broader macroeconomic outcomes. By focusing on the intersection of informality, entrepreneurship, and learning, this research connects with a significant body of literature on resource misallocation (Hsieh and Klenow (2009); Restuccia and Rogerson (2008); Guner et al. (2008); Garicano et al. (2016)), which has demonstrated that misallocation significantly reduces total factor productivity and output. While existing studies have underscored the importance of misallocation, the underlying mechanisms that drive it remain an area of active debate. Our paper addresses this gap by highlighting the role of firm learning as a critical, yet underexplored, determinant of misallocation. Specifically, we argue that learning dynamics within the informal sector, coupled with market frictions such as credit constraints, can exacerbate misallocation during the learning process.

By capturing the strategic nature of informality as a vehicle for learning, our framework provides a quantitative foundation for evaluating the effectiveness of formalization policies. In doing so, we contribute to the growing literature on policy effectiveness, particularly concerning interventions targeting the informal sector (Kaplan et al. (2011); Rocha et al. (2018); Bruhn and McKenzie (2014)). Our approach underscores the necessity of balancing regulatory requirements with opportunities for entrepreneurial experimentation, which are critical for fostering sustainable development in emerging economies. Through this exploration, we aim to inform policymakers about the com-

plex dynamics at play within the informal sector and the potential pathways to enhance economic growth and development.

Previous studies emphasize the importance of incorporating uncertainty and learning into models of firm behavior and entrepreneurship to explain the behaviors of private business owners and their macroeconomic implications (see, for example, Vereshchagina and Hopenhayn (2009); David et al. (2016); Chen et al. (2023); Kozlowski et al. (2020); and Arkolakis et al. (2018)). We distinguish our work from those studies in two key aspects. First, in addition to this literature highlighting how informational frictions affect the real economy by distorting firms' input decisions, we identify another channel through which information frictions create occupational misallocation by hindering high-potential individuals from becoming entrepreneurs. Second, we investigate the intricate interplay between informality and learning creates and study its effect on resource allocation and overall economic output.

3. Facts About Informality

This section characterizes the features of entrepreneurs and firms in both the formal and informal sectors, which serves as the motivation for our model. It also provides empirical foundations for the calibration of the quantitative model. We begin with an overview of the definitions for the key concepts and our primary data sources before presenting a series of key empirical findings.

3.1. Definitions and Data

We define informality in accordance with the existing literature: a firm is considered formal if it is registered with the tax authorities and possesses a tax identification number (e.g., Ulyssea (2018), Erosa et al. (2023)). Entrepreneurs are identified based on occupational status, specifically those classified as self-employed or employers. These definitions are consistently applied in both the theoretical model and the empirical analysis.

To conduct our empirical investigation, we rely on two Brazilian datasets. Brazil is an ideal setting for studying informality due to its widespread informal sector and the availability of high-quality data that comprehensively cover both informal and formal businesses. The richness of these datasets allows for an in-depth analysis of entrepreneurial characteristics and decision-making, capturing the intersection between household-level economic behavior (entrepreneurs) and firm-level dynamics (business operations). Importantly, Brazil's extensive data on informal firms—an aspect often underrepresented in other countries' datasets—enhances the robustness of our empirical approach.

Our primary dataset is the Brazilian National Household Sample Survey (PNAD), a repeated cross-sectional survey that provides detailed information on individuals' labor market outcomes, including income, occupation, formality status, and sociodemographic characteristics. UThis dataset enables us to identify both formal and informal sector entrepreneurs and analyze their attributes, such as age, gender, and educational background. Additionally, PNAD contains information related to their firms, including firm size and profit margins.

To ensure consistency across data sources, we focus on survey years from 2002 to 2009. The group of interviewees changes in each survey year. However, similar results are obtained using the data from every survey year. Hence, we compute the statistics by pooling across all years. Our sample includes working-age individuals (18–65 years old) with recorded employment. This selection yields approximately 100,000 observations per survey year. Notably, since 95% of individuals report having only one job, we restrict our analysis to information from the primary (full-time) occupation.

To gain deeper insights into the differences in firm behavior between the informal and formal sectors, we also leverage an additional dataset—the Urban Informal Economy Survey (ECINF), collected by the Brazilian Bureau of Statistics (IBGE) in 1997 and 2003. Our analysis focuses on the 2003 wave, which covers 48,701 entrepreneurs operating in urban areas across all Brazilian states. ECINF offers matched employer-employee data and provides granular firm-level information, including business age, employment size, investment levels, revenues, expenditures, and average profit margins. The survey is designed to be nationally representative for firms with up to five employees, making it particularly useful for studying the dynamics of small businesses in the informal economy. We primarily use this dataset for robustness checks, validating our findings from PNAD and ensuring the consistency of our empirical results.

3.2. Entrepreneurs in the Informal and Formal Sectors

Brazil exhibits both a high rate of entrepreneurship and a significant prevalence of informality among entrepreneurs, a pattern consistent with findings from other developing countries. In Brazil, entrepreneurs constitute 27% of the working population, yet only 30% of them operate businesses in the formal sector.

Existing literature using microdata from various countries has documented that informal entrepreneurs typically possess lower skill levels, often measured by educational attainment, and that informal firms tend to employ fewer workers and exhibit lower value-added per employee compared to their formal counterparts (La Porta and Shleifer (2014)). Our analysis of Brazilian data reveals similar patterns. The majority of entrepreneurs—particularly those in the informal sector—are self-employed individuals who manage and operate their businesses without hiring external labor. In fact, approximately 93% of informal firms have no employees. Additionally, only about 20% of informal entrepreneurs have completed at least a high school education, and informal firms, on average, exhibit lower productivity levels than their formal-sector counterparts.

Despite these significant differences, formal and informal entrepreneurs coexist within narrowly defined industries, even at the seven-digit industry classification level. This finding aligns with the results of Ulyssea (2020). Building on this evidence, our study takes a step further by exploring potential heterogeneity among informal entrepreneurs and their firms that may be masked by aggregate statistics. To do so, we classify individuals into income deciles and analyze the characteristics of formal and informal entrepreneurs within each income group.

Fact 1. Formal entrepreneurs are concentrated at the top income groups.

Figure 1 presents the population share of entrepreneurs across income deciles, distinguishing between formal (blue) and informal (red) entrepreneurs. The distribution reveals that formal entrepreneurs are largely absent from the lower half of the income distribution. Instead, they are predominantly concentrated at the upper end, with 62% of formal entrepreneurs belonging to the top three income deciles. Notably, 30% of formal entrepreneurs are in the top income decile.

Fact 2-1. Informal entrepreneurs are present across all income deciles, including the highest-income groups.

As shown in Figure 1, approximately 40% of households in the top income decile are entrepreneurs, and notably, 41% of them operate businesses in the informal sector. This finding contributes a novel insight to the literature, which has traditionally documented that informal entrepreneurs and their firms tend to generate lower profits. While a significant share of low-income entrepreneurs operate in the informal sector, our results indicate that informality is not exclusive to lower-income groups—some high-income entrepreneurs also choose to run businesses in the informal sector. This raises an important question: Who are these high-income informal entrepreneurs, and what types of businesses do they run?

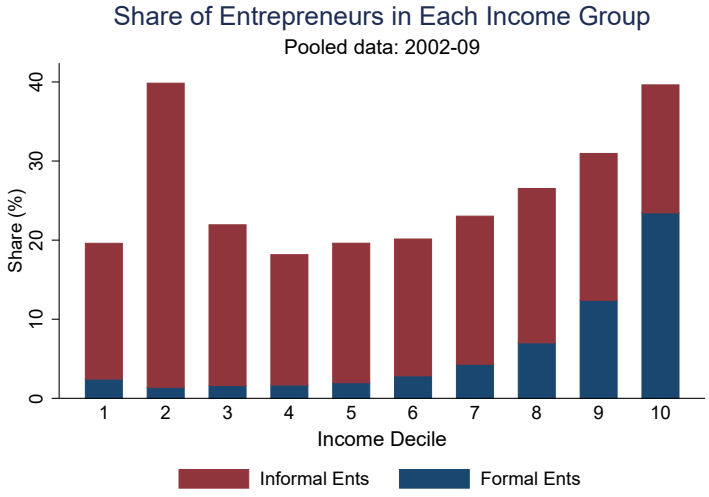


Figure 1. Share of Entrepreneurs in Each Income Decile

Notes: Data from PNAD 2002-2009. 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.

Fact 2-2. Informal entrepreneurs in the top income group are similar to their formal counterparts in terms of education level, firm productivity, and income.

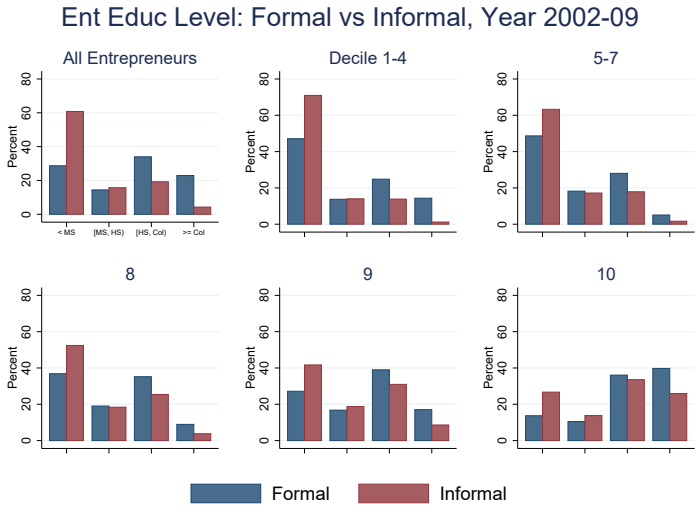


Figure 2. Entrepreneurs' Education Level in Each Income Group

Notes: Data from PNAD 2002-2009. This figure shows the share of informal entrepreneurs (red) and formal entrepreneurs (blue) across different education levels within each income group. 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 2 presents the distribution of formal (blue) and informal (red) entrepreneurs across four education levels: less than middle school, completed middle school but less than high school, completed high school but less than college, and college or above. The upper left panel displays this distribution for all entrepreneurs. Consistent with previous literature, informal entrepreneurs tend to have lower educational attainment—only about 25% have completed high school or higher, while 60% have not completed middle school. In contrast, nearly 60% of formal entrepreneurs hold at least a high school degree. However, when focusing on the top income decile (bottom right panel), a different pattern emerges. Informal entrepreneurs in this high-income group exhibit similar educational attainment to their formal counterparts. More than 60% of informal entrepreneurs in the top income decile have at least a high school degree, suggesting that high-income informal entrepreneurs are distinct from the broader informal sector and more comparable to formal entrepreneurs in terms of human capital.

To construct a proxy for productivity, we regress the log of profit on firm size and a set of industry dummies to purge inter-industry variation. Figure 3 presents the estimated productivity distributions for firms in the informal (red) and formal (blue) sectors. The left panel, which includes all firms, shows that the productivity distribution in the formal sector is shifted to the right, indicating that formal firms tend to be more productive on average than informal firms. However, when focusing on firms operated by entrepreneurs in the top income decile (right panel), the productivity distributions of formal and informal firms exhibit substantial overlap.

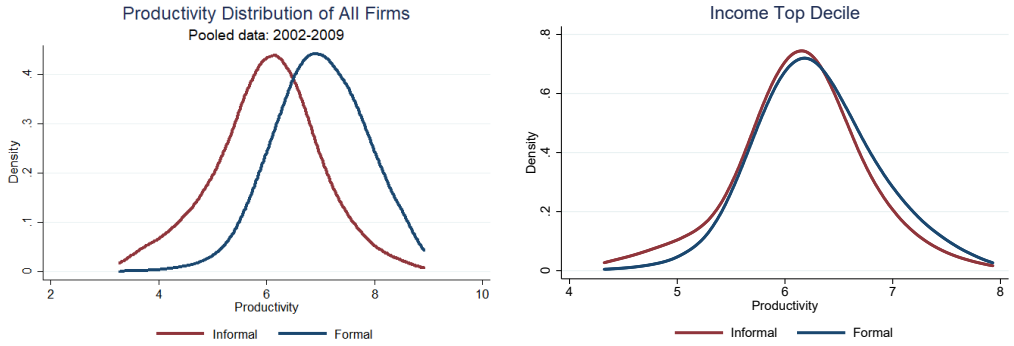


Figure 3. Firm Productivity

Notes: Data from PNAD 2002-2009. We regress the log of profit controlling for industry and firm size. The figure shows the densities of computed log-residuals for formal and informal firms. 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.

$log(income)$	All Ents	Decile 10
Formal (dummy)	0.30***	0.04***
Education (years)	0.06***	0.02***
Male (dummy)	0.40***	0.11***
Age (years)	0.03***	0.01***
Age squared	-0.00***	0.00
Race	Yes	Yes
State	Yes	Yes
Industry	Yes	Yes
Firm Size	Yes	Yes
Observations	193,647	28,344
R-squared	0.46	0.16

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

Table 1. Income Gap Between Formal and Informal Entrepreneurs

Notes: Data from PNAD 2002-2009. The table reports the income gap between formal and informal entrepreneurs when controlling for both individual and firm characteristics.

Next, we examine the income gap between formal and informal entrepreneurs. The previous literature documents that individuals in the informal sector have lower incomes on average than those in the formal sector. Our analysis confirms this pattern among entrepreneurs. After controlling for individual characteristics such as education, gender, age, and race, as well as business characteristics including location, industry, and firm size, we find that formal sector entrepreneurs earn, on average, 30% more than their informal sector counterparts (see Column 1 of Table 1). However, when restricting the sample to entrepreneurs in the top income decile, this income gap narrows significantly to just 4% (see Column 2 of Table 1).

These findings suggest that while informal entrepreneurs generally have lower education levels, operate less productive businesses, and have lower incomes, those in the top income decile resemble their formal-sector counterparts in key characteristics, including education, firm productivity, and income. This challenges the view that informality is purely a necessity-driven outcome. Instead, it suggests that for some entrepreneurs, operating informally may be a strategic decision rather than a constraint.

3.3. Transitions Between Informal and Formal Entrepreneurship

To examine whether informality is a strategic choice for some entrepreneurs, we analyze transitions between informal and formal entrepreneurship. This exercise provides deeper insights into the mechanisms driving informality. If entrepreneurs use the informal sector as a temporary state, we should observe some of them transitioning to the formal sector as they achieve success. In particular, we expect a significant share of high-income informal entrepreneurs to formalize their businesses over time. Conversely, if most high-income informal entrepreneurs remain informal, it raises the question of whether they do so due to other factors, such as tax avoidance, regulatory burdens, or barriers to formalization. Understanding these transitions allows us to assess whether the informal sector primarily serves as a stepping stone or a long-term equilibrium for certain entrepreneurs.

Since PNAD is a repeated cross-section survey, we turn to Continuous PNAD (PNADC), which follows individuals for five consecutive quarters, allowing us to observe sectoral transitions over a one-year period. We use PNADC 2012, the earliest available wave of the panel data, to compute transition rates from informal to formal entrepreneurship. This aligns with our main analysis, which relies on PNAD 2002–2009. As before, we restrict our sample to working-age individuals (18–65 years old) who are employed.

Fact 3. High-income informal sector entrepreneurs are more likely to transition to the formal sector than their lower-income counterparts.

Table 2 reports the transition rates of informal entrepreneurs into three categories: wage workers, informal entrepreneurs, and formal entrepreneurs, after one year. The first row presents data for all informal entrepreneurs. On average, 7.8% of informal entrepreneurs transition to the formal sector after one year, while the majority, about 72%, remain in the informal sector. The transition rate to the formal sector is even lower among low-income informal entrepreneurs, as shown in the second row of Table 2, which includes entrepreneurs in the bottom four income deciles. For this group, only 3.6% transition to the formal sector. However, high-income informal entrepreneurs exhibit a higher

likelihood of transitioning to the formal sector. According to the last row of Table 2, nearly 20% of informal entrepreneurs in income deciles 8 to 10 move to the formal sector. Among informal entrepreneurs in the top income decile, the transition rate is even higher, with 32% moving to the formal sector.

	t			
$t-1$	Worker	Informal Ent	Formal Ent	Total
Informal Ent (All)	19.97	72.24	7.80	100
Informal Ent (Decile 1-4)	20.56	75.84	3.61	100
Informal Ent (Decile 8-10)	14.45	66.11	19.44	100

Table 2. Transition Between Informal and Formal Entrepreneurship

Notes: Data from PNADC 2012.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 row reports transitions for high-income informal entrepreneurs (income deciles 8–10).

This pattern suggests that informality is not solely a sign of necessity but can be a strategic choice for some entrepreneurs. Informal entrepreneurs, especially those in the top income decile, appear to use the informal sector as a stepping stone to formalization, likely after achieving a certain level of success. The higher transition rates among more successful informal entrepreneurs support the idea that informality serves as a temporary state. Once their businesses become more established and profitable, these entrepreneurs are more likely to formalize their operations, potentially to access new opportunities, reduce regulatory risks, and expand their market reach.

4. Model

In the empirical section, we document three key facts that are novel to the literature: (i) More than one-third of entrepreneurs in the top income decile operate businesses in the informal sector, and these high-income informal entrepreneurs share similar characteristics with their formal counterparts in terms of education, productivity, and income; (ii) Entrepreneurs in the formal sector are predominantly concentrated in the higher income groups, suggesting they may bypass the typical entrepreneurial path of starting small and gradually scaling into the top income brackets; and (iii) High-income informal sector entrepreneurs are highly likely to transition to the formal sector. Motivated by these findings, we develop a theory of business learning and entrepreneurship in this section that can reconcile these empirical observations.

Specifically, the model incorporates uncertainty due to information friction, which prevents individuals from fully understanding their entrepreneurial capabilities. This uncertainty forces individuals to make a choice between wage work and entrepreneurship, without knowing their true potential as entrepreneurs. Entrepreneurs, therefore, must use their output in each period as a signal to update their beliefs about their abilities. In this

setting, the informal sector plays a crucial role by providing a low-cost platform for entrepreneurship experimentation. The absence of taxes and lower fixed operational costs in the informal sector allows entrepreneurs to test their business ideas and learn their business quality. This framework integrates learning dynamics with entrepreneurship and highlights the role of informality as a stepping stone in the entrepreneurial journey, providing new insights into the economic development process.

4.1. Demographics and Preferences

Building on Erosa (2001) and Franjo et al. (2022), we adopt an overlapping generations model that incorporates imperfect financial markets and distortionary taxation. Time is modeled as discrete, with each period representing one year. The economy is composed of a government, a representative corporate firm, and a continuum of households, all of whom are heterogeneous in their productivity levels.

Agents enter the economy at the age of 20 ($t = 1$), and make occupational decisions between wage employment and entrepreneurship throughout their working years. Upon reaching age 65 ($t = 46$), agents are mandatorily retired. From that period on, each household 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.

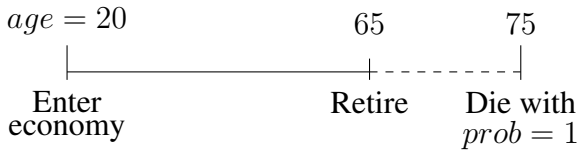


Figure 4. Life Cycle of the Model

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.

4.2. Endowments

Households are endowed with one unit of time each period. In addition, each household 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, households make occupation choices at each period during their working years, based on these two skills as well as other state variables.

Labor Market Productivity: Labor market productivity for individual i at age t, denoted by s_{it}, follows an AR(1) process, which is specified as follows:

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

Entrepreneurial 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 individual chooses to become an entrepreneur, their innate business quality q is fixed. This implies that the business idea's inherent quality is determined at the outset and remains constant once the entrepreneurial decision is made. A key assumption is that individuals do not observe the q they draw. As a result, households must decide between wage work and entrepreneurship without knowing the true quality of their business idea. If they opt for entrepreneurship, they also need to choose between the informal and formal sectors, based on their beliefs about the innate quality q of their business idea.

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

$$Q_t = q + e_t \tag{3}$$

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

4.3. 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:

$$f(k, n) = e^Q \Big(k^\alpha n^{1-\alpha} \Big)^\eta \tag{4}$$

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, s, as an input to production, in addition to hiring external labor, $n - s$. 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:

$$\pi^i = e^Q \Big(k^\alpha n^{1-\alpha} \Big)^\eta - (r + \delta)k - w(n - s) \cdot \mathbf{1}_{n>s} - w\tau(n) \tag{5}$$

where r_δ is the cost of renting capital (r is the rent rate and δ is the depreciation rate of capital), and $w(n - s)$ represents the wage bill for external labor. The indicator function

$\mathbf{1}_{n>s}$ ensures that wages are only paid to external labor when $n > s$, meaning there is at least some external labor hired. 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}. In every subsequent period, they earn the following profit:

$$\pi^f = e^Q \left(k^\alpha n^{1-\alpha}\right)^\eta - (r+\delta)k - w(1+\tau_{ss})(n-s) \cdot \mathbf{1}_{n>s} \tag{6}$$

Formal firms are registered with the tax authority, which means they must comply with all relevant taxes and regulations by paying a constant payroll tax, τ_{ss}. This tax is applied to the total wage bill for external labor.

Corporate Sector: Following **?**, 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:

$$F_c(K_c, N_c) = A_c K_c^\alpha N_c^{1-\alpha} \tag{7}$$

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.

4.4. Learning

Households do not initially 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, households have a prior belief about q that follows the population distribution of business quality, $\mathcal{N}(\mu_q, \sigma_q^2)$. When a household 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 household's belief about q follows the prior distribution $\mathcal{N}(\hat{\mu}_{q,n}, \hat{\sigma}^2_{q,n})$, 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:

$$\hat{\mu}_{q,n+1} = \frac{\hat{\sigma}^2_{q,n} Q_t + \sigma^2_e \hat{\mu}_{q,n}}{\hat{\sigma}^2_{q,t} + \sigma^2_e}$$
$$\hat{\sigma}^2_{q,n+1} = \frac{\hat{\sigma}^2_{q,n} \sigma^2_e}{(1+n)\sigma^2_q + \sigma^2_e} \tag{8}$$

In this process, $\hat{\mu}_{q,n+1}$ is the updated belief about q, and $\hat{\sigma}^2_{q,n+1}$ 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.

4.5. Asset Market and Borrowing Constraints

Households 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 households 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:

$$k \leq \lambda_f a \tag{9}$$

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

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

where $\lambda_i = 1$.

4.6. Government and Tax Systems

The government finances its expenditures, denoted by $\overline{G}$, through taxes on household consumption and wage payments by formal sector firms. Specifically, consumption is subject to a flat tax rate τ_c, while formal sector entrepreneurs pay a payroll tax at a flat rate τ_{ss} on the wage payment of external workers they hire.

4.7. The Household's Recursive Problem

The precise timing of decisions and events within a period is summarized in Figure 5. 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 a household 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 back to informality. To simplify the analysis, we assume that only entrepreneurs make sectoral choices—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 ws_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. This structure captures the dynamic learning process entrepreneurs undergo and the role of informality as a platform for experimentation and business discovery.

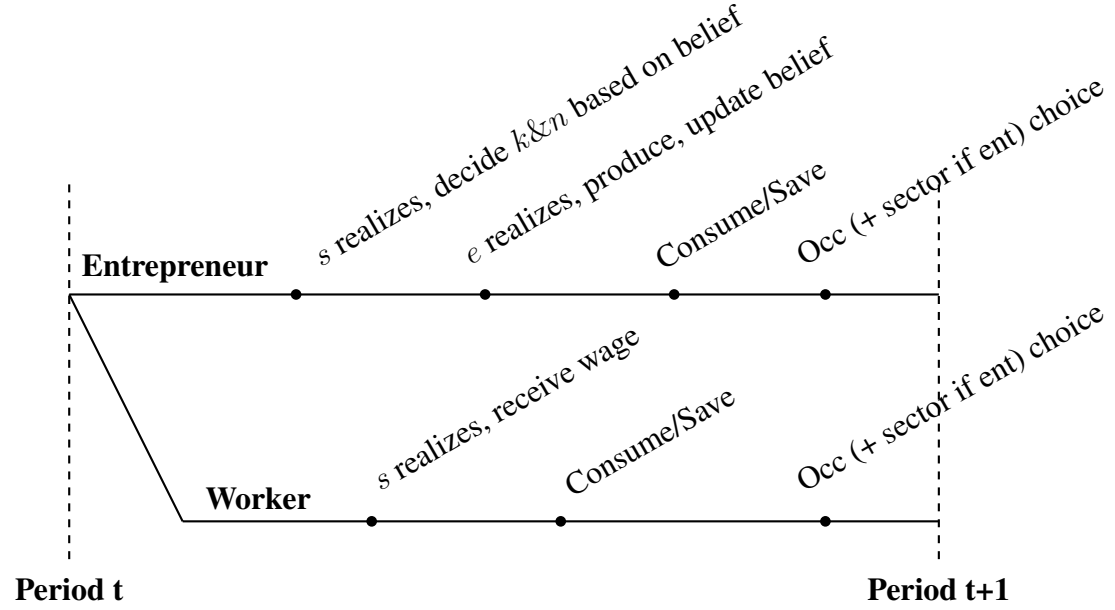


Figure 5. Timeline with One Period

In this model, the state variables for a working-age household are given by $x = (a, s, Q, \hat{\mu}_q, n, t)$, where a is the household'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 household'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 household's age. For retired households, the state space simplifies to just age and asset holdings, given by (t, a).

Old Retirees: From $t = 46$ to 55 (physical age 65 to 74), households are retired and survive to the next period with a probability that is decreasing in their age. 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

$$\begin{aligned} V^r(a,t) &= \max_{c,a'} \; u(c) + \beta(1-\Omega_{t+1})V^r(a',t+1) \\ \text{s.t.} \quad & (1-\tau_c)c + a' = (1+r)a \\ & a' \geq 0 \end{aligned} \tag{11}$$

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

$$\begin{aligned} V^w(a,s,t) &= \max_{c,a'} \; u(c) + \beta \cdot \mathbf{1}_{t=45} \cdot (1-\Omega_{t+1})V^r(a',t+1) \\ &+ \beta(1-\mathbf{1}_{t=45}) \max \Big\{ \mathbb{E}[V^w(a',s',t+1)], \mathbb{E}[V^i(a',s',Q',\mu_q,0,t+1)], \\ &\qquad \mathbb{E}[V^f(a'-C^f_{entry},s',Q',\mu_q,0,t+1)] \Big\} \\ \text{s.t.} \quad & (1-\tau_c)c + a' = ws + (1+r)a \\ & a' \geq 0 \end{aligned} \tag{12}$$

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 payroll taxes for their production, but, in avoiding taxation, they face distortions in their production decisions and incur an additional cost, $\tau(n)$. If an informal entrepreneur operates as self-employed, the wage bill is zero. If they hire external labor, the total wage bill is given by $w(n-s)$. The decision problem of an

informal sector entrepreneur is formulated as follows:

$$V^i(a,s,Q,\hat{\mu}_q,n,t) = \max_{k,n,c,a'} \; u(c) + \beta \cdot \mathbf{1}_{t=45} \cdot (1-\Omega_{t+1}) V^r(a',t+1) \tag{13}$$
$$+ \beta(1-\mathbf{1}_{t=45}) \max \Big\{ \mathbb{E}[V^w(a',s',t+1)], \mathbb{E}[V^i(a',s',Q',\hat{\mu}'_q,n+1,t+1)],$$
$$\mathbb{E}[V^f(a'-C^f_{entry},s',Q',\hat{\mu}'_q,n+1,t+1)] \Big\}$$
$$\text{s.t.} \quad y = e^Q \Big(k^\alpha n^{1-\alpha}\Big)^\eta - (r+\delta)k - w(n-s) - w\tau(n)$$
$$(1-\tau_c)c + a' = y + (1+r)a$$
$$k \le a, \quad a' \ge 0$$

The recursive problem of formal sector entrepreneurs closely resembles that of informal sector entrepreneurs, with a few key differences. Formal entrepreneurs are subject to a payroll tax on external labor but do not incur the additional production distortion faced by informal entrepreneurs. Additionally, they have access to external financing, allowing them to borrow capital from financial intermediaries. Most importantly, formal entrepreneurs have the choice to either continue operating their business or shut it down and return to wage employment. However, once they have entered the formal sector, they cannot convert their business back to the informal sector.

$$V^f(a,s,Q,\hat{\mu}_q,n,t) = \max_{k,n,c,a'} \; u(c) + \beta \cdot \mathbf{1}_{t=45} \cdot (1-\Omega_{t+1}) V^r(a',t+1) \tag{14}$$
$$+ \beta(1-\mathbf{1}_{t=45}) \max \Big\{ \mathbb{E}[V^w(a',s',t+1)], \mathbb{E}[V^f(a',s',Q',\hat{\mu}'_q,n+1,t+1)] \Big\}$$
$$\text{s.t.} \quad y = e^Q \Big(k^\alpha n^{1-\alpha}\Big)^\eta - (r+\delta)k - (1+\tau_{ss})w(n-s)$$
$$(1-\tau_c)c + a' = y + (1+r)a$$
$$k \le \lambda_f a, \quad a' \ge 0$$

4.8. Stationary Equilibrium

I focus on a stationary equilibrium concept where factor prices and distribution are constant over time.

Define $x = (a,s,Q,\hat{\mu}_q,n,t)$ as the state vector. Given a tax structure $\{\tau_c,\tau_{ss}\}$, a stationary equilibrium is defined by the household value function, their decision rules $c(x)$, $a'(x)$, $n(x)$, $k(x)$, corporate sector choices K_c and N_c, factor prices $\{r,w\}$, and a distribution over households $\Lambda(x)$, such that, given prices and government tax schedules:

1. policy functions $c(x)$, $a'(x)$, $n(x)$, $k(x)$ maximize household's value function;
2. the corporate firm maximizes profit
3. the government budget is balanced
4. the goods market clears

$$Y = \int c(x) d\Lambda(x) + K' - (1-\delta)K + \overline{G} + \int C_{entry} \cdot \mathbf{1}_{w,i\to f} \cdot \Lambda(x) \tag{15}$$

5. the labor market clears

$$N_f + N_i + N_c = \int s(x) \cdot \mathbf{1}_{occup=w} \cdot \Lambda(x) \tag{16}$$

6. the capital market clears

$$K_f + K_i + K_c = \int k(x)\Lambda(x) \tag{17}$$

7. the distribution of household over the state variables, $\Lambda(x)$, is stationary.

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 3. In the second step, the remaining 5 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 4.

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[1]. 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).

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

Production Technologies: Corporate productivity A_c is normalized to be one. The capital share of corporate firm's production function α is set to be 0.406 as in Allub and Erosa (2019) and Franjo et al. (2022). For simplicity, we make the value of the capital share of the entrepreneurial sector equal to that of the corporate sector. 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 0.06, which is a standard value in the literature, as in Buera and Shin (2013). Formal sector entrepreneurs can borrow up to 37.8 percent of their assets, allowing for a maximum investment of $\lambda_f = 1.378$ times their assets, as estimated by Franjo et al. (2022). In contrast, informal sector entrepreneurs must rely solely on their own assets to finance capital inputs, implying $\lambda_i = 1$ (Franjo et al. (2022)).

The informal sector firms face a labor distortion, denoted by $\tau(n)$, which increases and is convex in firm size. Following Ulyssea (2018), we assume the functional form $\tau(n) = \frac{n^2}{b}$, where $b > 0$. The final production-related parameter is the one-time fixed entry cost to the formal sector, C^f_{entry}. Both b and C^f_{entry} will be calibrated internally. A more detailed discussion of their calibration will be provided later, when we address the identification strategy for each calibrated parameter.

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 payroll tax rate is set at its statutory value of $\tau_{ss} = 0.29$.

Parameter	Description	Source	Value
Preference			
γ	Relative risk aversion	Buera and Shin (2013)	1.5
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	Buera and Shin (2013)	0.06
λ_i	Informal sec borrowing constraint	Franjo et al. (2022)	1
λ_f	Formal sec borrowing constraint	Franjo et al. (2022)	1.378
Taxes			
τ_c	Consumption tax	Jung and Tran (2012)	0.15
τ_{ss}	Payroll tax	Statutory Value	0.29

Table 3. Parameters Calibrated Outside of the Model

5.2. Parameter Values Calibrated Within the Model

In addition to the two parameters related to entrepreneurial production–namely, b, which governs the labor distortion for informal sector firms, and the one-time fixed entry cost to the formal sector, C^f_{enty}–there are three parameters that govern the business quality process, which will also be calibrated internally. Specifically, the overall business productivity Q_t consists of two components: the fixed innate business quality q, drawn from a normal distribution $\mathcal{N}(\mu_q, \sigma_q^2)$, with this distribution also serving as the initial belief for startup entrepreneurs, and a transitory shock e_t which is an $i.i.d$ draw from a normal distribution $\mathcal{N}(0, \sigma_e^2)$ each period. Therefore, the parameters μ_q, σ_q^2, and σ_e^2 govern the Q process.

As μ_q is the initial belief about the innate business quality, households decide if they want to start a business based on this belief. Other factors being equal, a higher belief will encourage more households to pursue entrepreneurship. To calibrate μ_q, we target the entrepreneur population rate, which is estimated to be 0.32 using 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, q becomes the primary determinant of the firm's size, as e_t is just a transitory shock and entrepreneurs make their input decision before the realization of e_t. 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 match the share of formal firms with at least 50 employees, a statistic estimated by Ulyssea (2018) to be 0.02 using RAIS data, an administrative dataset that includes the universe of formal firms.

σ_e governs the dispersion of the transitory shock. For firms with an accurate belief about their innate business quality, fluctuations in their profit stream are primarily driven by this transitory shock, since entrepreneurs make their input decisions based on their belief about q before the realization of e_t. Therefore, we calibrate σ_e by matching the formal sector entrepreneurs' profit autocorrelation after one year, which is estimated to be 0.72 using PNADC. We use the profit data of formal sector entrepreneurs as the identification strategy. The underlying assumption is that entrepreneurs use the informal sector as a low-cost platform to learn about their business quality. As their businesses become more established and profitable, these entrepreneurs are more likely to formalize their operations. Therefore, formal sector entrepreneurs are those with a more accurate belief about their business quality, and, as a result, the fluctuation in their profit stream is primarily driven by the transitory shock.

The entry cost C^f_{entry} influences the share of formal businesses in the economy. We calibrate this parameter by targeting the fraction of formal businesses, which is estimated to be 0.3 using PNAD data. The final parameter, b, governs the cost of informality. The distortion $\tau(n)$ is inversely related to b; as b decreases, it becomes more costly for informal sector entrepreneurs to employ labor, leading them to prefer smaller business sizes. We calibrate b by targeting the share of informal firms with zero employees, which is estimated to be 0.93 using PNAD data.

5.3. Calibration Results

The model economy performs well in matching the targeted moments. Table 4 presents the calibration results, including parameter values, targeted moments, and comparisons between model and data moments.

Parameter	Value	Targeted Moment	Data	Model
μ_q	0.07	Population share of ents (PNAD)	0.27	0.27
σ_q	0.66	Share of formal firms with $\geq$ 50 employees (RAIS)	0.03	0.03
σ_e	0.40	Formal ents' profit autocorrelation after 1 yr (PNADC)	0.72	0.74
C^f_{entry}	2.40	Share of formal firms (PNAD)	0.30	0.31
b	3.48	Share of informal firms with emp $\leq$ 2 (PNAD)	0.96	0.98

Table 4. Parameters Calibrated (Jointly) Inside the Model

The model successfully replicates the high rate of entrepreneurship observed in Brazil, where entrepreneurs make up 32% of the labor force, compared to the model's prediction of 37%. To generate this high entrepreneurship rate, the model requires the initial belief about innate business quality, μ_q, to be set at 0.07. Additionally, the model captures the right tail of the size distribution within the formal sector, although it slightly overestimates it. While the fraction of formal firms with at least 50 employees is 0.02 in the data, the model predicts 0.08. However, Figure 6 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 0.4.

Regarding the two parameters related to entrepreneurial production, the one-time fixed entry cost to the formal sector is calibrated to 2.4 to match the observed fraction of formal firms, which is 0.3. For comparison, the annual wage income of a worker with the lowest productivity is 1.15. This implies that the entry cost of 2.4 is approximately twice the annual wage income of the lowest-earning workers in the economy. Additionally, by setting $b = 3.48$, the model accounts for the fact that most informal sector firms do not hire external workers. In the model, the fraction of self-employed informal entrepreneurs is 0.97, compared to 0.93 in the data.

5.4. Model Performance in Non-targeted Dimensions

We also assess how the model performs in comparison to selected non-targeted moments, with results reported in Table 5. The most important finding is that our model of business learning, entrepreneurship, and sectoral choice–motivated by the three empirical facts we document–successfully reproduces these patterns despite not using them as calibration targets.

Moment	Source	Data	Model
Frac. of formal ents in the income top decile (***Fact 1***)	PNAD	0.30	0.37
Frac. of inc. top decile ents that are informal (***Fact 2***)	PNAD	0.41	0.43
Frac. of informal ents that switch to formal	PNADC	0.08	0.08
Frac. of inc. decile 8-10 informal ents that switch to formal ((***Fact 3***))	PNADC	0.19	0.18
Credit-to-GDP ratio	Franjo et al. (2022)	0.42	0.32
Wealth Gini	Davies et al. (2011)	0.78	0.80
Ave age of informal / formal ents	PNAD	41/44	39/44

Table 5. Model Performance: Untargeted Moments

First, the model captures the concentration of formal entrepreneurs in higher income groups. In the data, 30% of formal entrepreneurs belong to the top income decile, while our model produces a similar figure of 37%. Moreover, the model correctly predicts that formal entrepreneurs are largely absent from lower income groups. Second, the model closely matches the fraction of top-income decile entrepreneurs who operate informally: 0.41 in the data versus 0.43 in our model. Third, the model accurately replicates the transition dynamics of high-income informal sector entrepreneurs. In the PNADC data, the one-year transition rate from informality to formality is 0.08 overall but rises to 0.19 for informal sector entrepreneurs in income deciles 8-10. Our model closely mirrors these figures, yielding transition rates of 0.08, and 0.18, respectively. These results demonstrate that the model not only captures the key patterns of informality and entrepreneurship but also provides a useful framework for analyzing the underlying mechanisms driving these outcomes.

In our calibration, we set the collateral constraint for formal sector entrepreneurs to 1.378, following Franjo et al. (2022). Our model produces a credit-to-GDP ratio of 0.32, which is reasonably close to the data estimate of 0.42. This parameter choice suggests that financial frictions are more severe in the informal sector than in the formal sector, a finding consistent with previous studies such as Franjo et al. (2022) and Erosa et al. (2023). The model also aligns well with data on wealth inequality. It generates a wealth Gini coefficient of 0.80, closely matching the wealth Gini coefficient observed in the data, which is 0.78. Additionally, the model captures the pattern that formal sector entrepreneurs tend to be older than their informal sector counterparts. According to PNAD data, the average age of informal sector entrepreneurs is 41 years, while that of formal sector entrepreneurs is 44 years. Our model closely replicates this pattern, yielding average ages of 39 and 44, respectively. Finally, we examine the firm size distribution in the formal sector. In Figure 6, we compare our model's firm size distribution with that of Ulyssea (2018). The model developed by Ulyssea (2018) provides a good fit for the entire firm size distribution in the formal sector, closely matching estimates based on RAIS, an administrative dataset. Since we do not have direct access to RAIS, we compare our model results to those from Ulyssea (2018). Notably, even though our calibration explicitly targets only the right tail of the distribution, our model successfully captures the entire firm size distribution.

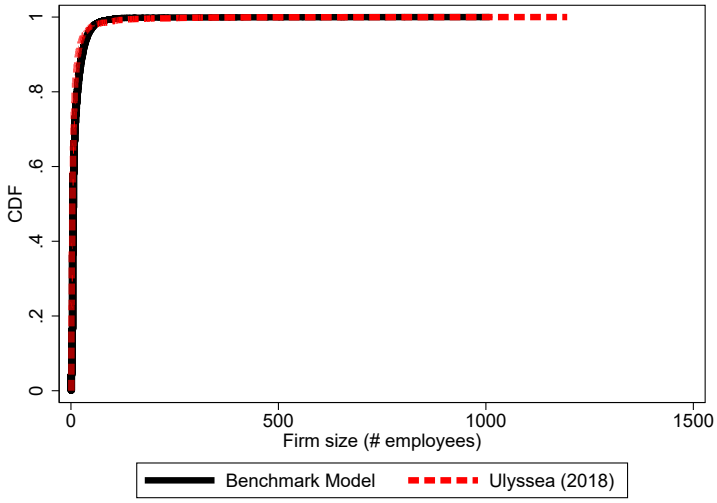


Figure 6. Firm Size Distribution in the Formal Sector: Model vs. Ulyssea (2018)

Notes: The model developed by Ulyssea (2018) provides a good fit for the entire firm size distribution in the formal sector, as estimated using RAIS data—an administrative dataset collected by the Ministry of Labor. Since we do not have direct access to RAIS, we instead compare our model's results to those from Ulyssea (2018).

6. Policy Counterfactuals

In this section, we examine the impact of formalization policies at the firm level and their broader macroeconomic implications. Specifically, we consider a policy that increases the cost of operating informally, which can be interpreted as a result of increased government monitoring. In the model, this translates into a lower value of the parameter b. We focus on this formalization policy because prior research has shown that stricter enforcement on the extensive margin is highly effective in reducing informality while boosting productivity and tax revenues (see, for example, Ulyssea (2018) and Erosa et al. (2023))[2].

Following the approach in Ulyssea (2018), which reduces b by a factor of 0.15, and Erosa et al. (2023), which applies a factor of 0.2, we adopt a similar strategy. In our policy counterfactual, we reduce b by multiplying it by 0.2. All other parameters remain fixed at their benchmark values, including the level of government expenditure, G. Any additional tax revenue generated from this formalization policy is redistributed to households through a lump-sum transfer, denoted by T, ensuring that the government budget remains balanced. If T turns negative, it effectively acts as a lump-sum tax to offset any potential decline in tax revenue. As highlighted by Ulyssea (2018), this experiment relies on two

[2]Ulyssea (2018) find that increasing the cost of being an informal firm can reduce informality in terms of the share of firms, workers and output. Additionally, this formalization policy improves total factor productivity (TFP) by 8.3%, increases total output by 3.2%, and raises total tax revenue by 22.2%. Similarly, Erosa et al. (2023) show that enhancing government monitoring of the informal sector yields comparable results. Their findings indicate that stricter enforcement reduces informality while boosting TFP, output, and tax revenues by 6.6%, 10.8%, and 33.2%, respectively.

key assumptions: (i) all tax revenues are fully rebated to households without any losses, and (ii) there are no costs associated with increased enforcement, despite the likelihood that real-world implementation would be costly, given the challenge of monitoring a large number of small firms.

All policy counterfactuals are conducted in a general equilibrium setting, where prices (r and w) are recalibrated to clear factor markets. Our analysis focuses on steady-state comparisons to evaluate the long-run effects of the policy.

	Baseline	Greater Monitoring	Δ
Informality			
Informal ents (mass)	1.00	0.23	-76.66%
Informal ents (share)	0.70	0.51	-19 p.p.
Informal output (share)	0.12	0.03	-9 p.p
Aggregate Economy			
Output	100	67.13	-32.87%
Measured ent TFP	9.73	8.83	-9.24%
Tax revenues	100	54.41	-45.59%
Welfare	100	72.02	-27.98%
Occupation Misallocation			
Num. of ents (mass)	1.00	0.32	-68.39%
Num. of formal ents (mass)	1.00	0.50	-49.58%
Num. of HHs with q_{max} (mass)	1.00	0.74	-26.44%
Frac. of q_{max} HHs who are ents	0.66	0.31	-35 p.p.

Table 6. 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 reducing the parameter b by 20%. The policy counterfactual is conducted in partial equilibrium, holding factor prices at their benchmark values. (i) 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. (ii) The welfare measure is given by $w\overline{L} + \Pi +$ Tax Revenue, where Π denotes total profits in the economy net of total entry costs.

Table 6 presents the results of increasing the cost for informal sector entrepreneurs. Consistent with existing literature, we find that this formalization policy significantly reduces informality. The number of informal entrepreneurs declines by nearly 77%, and the share of informal firms in the economy drops from 70% to 51% in the new steady state. Similarly, the share of informal sector output decreases by 9 percentage points, from 12% to just 3%. Despite its effectiveness in reducing informality, this policy imposes substantial economic costs in our model. Aggregate output declines by approximately 33%, measured entrepreneurial TFP falls by over 9%, total tax revenue drops by around 46%, and welfare decreases by nearly 28%. These losses are primarily driven by occupation misallocation.

In our model, due to information frictions, households do not observe their innate business quality upfront. They learn about it only through production. The informal sector serves as a low-cost platform for business experimentation. However, increased government monitoring reduces the value of informal entrepreneurship. Given their initial beliefs about business quality, households often perceive the value of starting a formal sector business as low due to entry costs and payroll taxes. The incentive to operate formally only increases once households have a stronger belief that their business quality is sufficiently high. As a result, this policy effectively raises barriers to business experimentation, leading to a 68% decline in the number of entrepreneurs. The share of entrepreneurs in the population falls from 37% to just 12%, and the number of formal sector entrepreneurs declines by 49.6%.

Ideally, all households with the highest business quality, q_{ma}, should start a firm rather than remain workers. However, as shown in Table 6, increased government monitoring exacerbates occupational misallocation by discouraging high-productivity individuals from becoming entrepreneurs. In the new steady state, the number of households with the highest business quality decreases by 26%. This occurs because, in each period, non-entrepreneurial working-age households draw a new business idea. They can only retain their business quality if they start a firm. In the baseline economy, the relatively low cost of informal sector entrepreneurship allows more households to enter the informal sector, experiment, and retain their high q once discovered. However, with increased monitoring and higher costs, fewer households choose to become entrepreneurs, reducing opportunities for business experimentation. As a result, even those who initially draw a high q cannot retain it, as they do not start a business and must redraw a new business idea in the next period.This effect is further illustrated by the fraction of households with q_{max} who become entrepreneurs: in the baseline economy, 66% of those with the highest business quality start a firm, but in the new steady state, this fraction drops to just 31%. This suggests that raising the cost of informal entrepreneurship leads to greater occupational misallocation, ultimately reducing productivity and output.

Another important aspect to consider is the decline in total tax revenues, which decrease by 28%. A significant portion of this decline comes from a 61.6% reduction in payroll tax revenue. This drop is driven by two key factors. First, there are fewer formal sector firms in the new steady state. Second, in the baseline economy, most entrepreneurs undergo their learning process in the informal sector and transition to the formal sector only after their business is well-established. As a result, formal sector entrepreneurs in the baseline economy tend to operate at a larger scale. However, with higher costs associated with informal entrepreneurship, more entrepreneurs enter the formal sector earlier, before their businesses have matured. This shift leads to a smaller average business size in the formal sector, further reducing payroll tax revenue.

More importantly, the decline in tax revenues also stems from a 38% decrease in consumption tax revenue. In the new steady state, with lower total tax revenues, the government must impose a lump-sum tax of 0.23, which amounts to roughly 20% of the annual wage income of the lowest-earning workers. This lump-sum tax reduces household income, leading to lower consumption. Additionally, some households that could have started a business and thus followed a higher-earning life-cycle trajectory, are now

deterred by the higher cost of entrepreneurship and remain as workers. Since their worker earnings are lower than what they could have earned as entrepreneurs, this further contributes to the decline in overall consumption in the new steady state.

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 an occupational choice model. In our model, entrepreneurs face uncertainty about their business quality and must learn through production. The informal sector provides a low-cost environment for this learning process, allowing high-potential businesses to emerge and transition to the formal sector over time.

We evaluate the macroeconomic effects of a policy that increases the cost of informal entrepreneurship, interpreted as intensified government monitoring and enforcement. Consistent with the existing literature, we find that this policy effectively reduces informality, lowering the number of informal entrepreneurs and reducing informal sector output. However, by restricting access to business experimentation, the policy also introduces substantial economic distortions. Total entrepreneurship declines by 68%, the number of formal sector entrepreneurs drops by 49.6%, and occupational misallocation worsens as fewer high-productivity individuals enter entrepreneurship. These misallocations lead to significant aggregate consequences: output falls by 33%, entrepreneurial productivity declines by 9%, and total tax revenue shrinks by 46%. The government is forced to impose a lump-sum tax to offset the revenue loss, further depressing consumption and welfare.

By explicitly modeling learning dynamics, we show that informality plays a crucial role in entrepreneurial selection and business dynamism. Policies that raise barriers to informal business operations may unintentionally suppress the emergence of productive firms, leading to long-run economic costs. Our results suggest that policies aimed at reducing informality should be carefully designed to preserve the benefits of business experimentation while addressing the constraints associated with informal sector entrepreneurship.

More broadly, this paper contributes to the literature on informality and entrepreneurship by providing a framework that links business learning to sectoral choices and macroeconomic outcomes. Our approach offers a new perspective on the informal sector—not simply as a distortion, but as a mechanism that enables the discovery of high-productivity entrepreneurs who might otherwise never enter the formal economy. Future research could explore alternative policy interventions, such as targeted credit access or phased formalization strategies, that balance the benefits of business learning with the efficiency gains of formality.

By shedding light on the interaction between informality, learning, and entrepreneurial selection, this paper provides new insights into the broader debate on economic development and policy design in economies with large informal sectors.

References

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

Alonso-Ortiz, J. and Leal, J. (2018). Cross-subsidies, and the elasticity of informality to social expenditures: The case of mexico's seguro popular. *Review of Income and Wealth*, 64(2):482–512.

Arkolakis, C., Papageorgiou, T., and Timoshenko, O. A. (2018). Firm learning and growth. *Review of Economic Dynamics*, 27:146–168.

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.

Busso, M., Fazio, M. V., and Algazi, S. L. (2012). (in) formal and (un) productive: the productivity costs of excessive informality in mexico. Technical report, IDB working paper series.

Cagetti, M. and De Nardi, M. (2006). Entrepreneurship, frictions, and wealth. *Journal of political Economy*, 114(5):835–870.

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.

Davies, J. B., Sandström, S., Shorrocks, A., and Wolff, E. N. (2011). The level and distribution of global household wealth. *The Economic Journal*, 121(551):223–254.

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.

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.

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

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.

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.

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.

Kozlowski, J., Veldkamp, L., and Venkateswaran, V. (2020). The tail that wags the economy: Beliefs and persistent stagnation. *Journal of Political Economy*, 128(8):2839–2879.

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

Levy, S. (2008). Social policy, informality, and economic growth in mexico.

Quadrini, V. (2000). Entrepreneurship, saving, and social mobility. *Review of economic dynamics*, 3(1):1–40.

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.

Vereshchagina, G. and Hopenhayn, H. A. (2009). Risk taking by entrepreneurs. *American Economic Review*, 99(5):1808–1830.