Informality and the Life Cycle of Plants*

Furkan Sarıkaya† M. Nazım Tamkoç‡ Jesica Torres§

November 2024

Abstract

This paper documents the life cycle of formal and informal plants using five waves of the Mexican establishment census. Formal plants begin operations with three times more workers than informal plants and demonstrate faster growth rates. Throughout their life cycle, formal establishments more than double their size, while informal plants increase their size by only 77%. A general equilibrium model is developed to quantify the aggregate economic losses stemming from these growth rate disparities. In the model, plants grow through productivity investments, and informality emerges from incomplete enforcement. In equilibrium, informal plants exhibit flatter life cycle profiles to avoid detection and taxation. Model parameters are calibrated to match key properties of plant size distribution and the life cycle of plants in Mexico. Quantitative results indicate that a revenue-neutral full enforcement increases aggregate output and the overall growth rate by sixteen and twenty-five percent relative to the benchmark, respectively.

JEL classification: E23, J24, L25, O41, O33
Keywords: Informality, Life-cycle, Development, Productivity, Distortions

*We would like to thank Edgar Avalos for his superb research assistance. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. Tamkoç and Torres thank DECRG for providing Research Support Budget. INEGI has reviewed our results to ensure that no confidential plant-level information is disclosed.

†Arizona State University

‡The World Bank. Corresponding author: M. Nazım Tamkoç, `mtamkoc@worldbank.org`.

§The World Bank

1. Introduction

Hsieh and Klenow (2014) have documented that plants in the U.S. grow more over their life cycle compared to plants in Mexico and India.[1,2] Less developed countries also face widespread informality.[3] This paper argues that informality, arising from incomplete enforcement, distorts incentives for growth over the life cycle.[4] Since tax compliance increases with plant size, informal plants may optimally choose to keep operating at smaller scales to escape detection. Hence, we study the role of informality in explaining the life cycle of plants. We ask how aggregate economic outcomes such as output and plant-size distribution, as well as the life cycle of plants, would change if informality were reduced by improving enforcement or reducing the burden of formality.

One of the key challenges in studying how informality distorts plant life cycles is the limited availability of plant-level panel data containing information on registration status. This study overcomes this challenge by analyzing Mexico's census plant-level data from Instituto Nacional de Estadística, Geografía e Informática (INEGI). The data tracks plants over 20 years, from 1998 to 2018, and includes information about establishments' registration with different tax authorities. Therefore, unlike Hsieh and Klenow (2014), this allows for analysis of a balanced panel of registered and unregistered plants.

In our analysis, an informal establishment is a plant which is not registered with either the central tax authority or the social security administration. In contrast, we refer to plants as formal if they are registered with either the central tax authority or the social security administration. We first show that young formal plants are bigger in terms of employment than young informal plants. Formal establishments in Mexico start operations with 6.1 workers on average whereas their informal counterparts report only 1.9 workers on on average. Second, formal plants grow faster over their life cycle compared to informal plants. While formal plants grow by a 136% during their life cycle, informal plants are only 1.77 times bigger when old compared to young informal plants.

In order to quantify the losses of informality from distortions to the life cycle growth

[1]We use the terms establishment and plant interchangeably throughout the paper.

[2]Eslava et al. (2022) shows that Colombian plants grow less than plants in the U.S.

[3]ILO estimates 57.4% and 89.1% of the employment as informal employment in Mexico and India, respectively whereas 77% of plants operate informally in Mexico according to INEGI census data.

[4]Compliance with taxes varies with plant size: the effective tax rate is higher among large establishments compared to smaller plants (See López and Torres (2020) for Mexico).

of plants, we develop a one-sector growth model where both formal and informal plants grow by investing in their productivity. Our model is based on Guner et al. (2018) (GPV in what follows) which in turn builds on the Lucas (1978). In the model, agents choose to be workers and entrepreneurs. Our main innovation to their setup is introducing a proportional tax on output that entrepreneurs can escape by operating informally due to incomplete enforcement, in the spirit of Leal (2014). There are two enforcement mechanism in the model. First, tax officials are not able to enforce taxes to all plants. Agents, at the beginning of their life-cycle, face a tax official with some probability. Second, regardless of whether agents are assigned to an official at the beginning of their life-cycle, informal entrepreneurs are caught if they use capital more than a certain amount. Entrepreneurs must therefore choose whether to operate informally or formally as well as when to transition into formality if they start operating informally at the beginning of their life-cycle.

In equilibrium, smaller and relatively unproductive plants operate informally, while larger plants and those assigned to tax officials early in their life cycle operate formally. Informal plants with higher productivity eventually transition to formal operation as they invest and grow. Consequently, less productive informal plants grow more slowly than formal plants to avoid tax enforcement.

We calibrate the model parameters to match key properties of the plant-size distribution and the growth rates of both formal and informal plants in Mexico. The model is able to successfully match the mean sizes of formal and informal plants as well as the concentration of employment in large plants. In addition, the model is able to capture the markedly different growth rates of formal and informal plants over their life cycle.

We first document the effects of changes in the burden of formality, i.e. the tax rate. As the tax rate declines, the formal managers increases their demand for inputs and marginal informal plants decides to operate formally to enjoy higher output and growth potential. This reduces the fraction of informal plant numbers and increases productivity investment, resulting in larger plants and higher aggregate output and growth. Lower taxes also increases the fraction of informal plants transitioning into formality along their life-cycle. Quantitatively, When the tax rate halves, the fraction of transitioning plants rises from 1.9% to 5.8%. This is another factor that contributes to higher growth rate of plants along their life-cycle on average because of the level effect. In other words, the transitioning plants starts their life-cycle small and informal; and they don't have any limit to grow when they become formal. So their growth rates are higher compared to other plants at the similar productivity levels.

Another formalization policy studied in the literature is through stricter enforcement. We experiment the effects of improving enforcement policies in revenue neutral set-up on the aggregate variables and the life-cycle of plants on average. The mean size and aggregate output increases by 63.5% and 16.4% under revenue-neutral full enforcement scenario. When taxes are fully enforced such that there is no option of operating informally, plants have no incentive to stay small. Therefore, overall plant growth rate increases by 25.2% and the entrepreneurial quality increases by 68.1%.

Background Our paper contributes to the informality and misallocation literature. The informality literature, as explored by De Soto (1989), La Porta and Shleifer (2008), Maloney (2004) and Loayza (2016) provides a comprehensive understanding of the economic implications of informal sectors.[5]. Our research is closely related to the works of Leal (2014), Meghir et al. (2015), Ulyssea (2018), and Tamkoç (2024b). We contribute to this body of work by demonstrating that informal establishments do not growth as much as formal establishments do throughout their life-cycle and by developing a model to quantify the impact of informality on aggregate output.

In addition, our paper engages with the misallocation literature, including seminal works by Hsieh and Klenow (2009), Restuccia and Rogerson (2008) and Bartelsman et al. (2013).[6] We specifically draw on literature examining the growth of plants over their life cycle, such as Hsieh and Klenow (2014), Guner et al. (2018), Garcia-Macia et al. (2019) and Eslava et al. (2022). Our unique contribution lies in constructing a model where formal and informal firms coexist and grow over their life cycle, thereby providing new insights into the dynamics of plant growth and the role of informality in economic development.

2. Establishment-level data

Our quantitative exercises exploit data from the Mexican establishment census.[7] The census is administered by Mexico's statistics agency INEGI and targets the universe of establishments in non-rural areas.[8] Enumerators conduct door-to-door visits and

[5]For a detailed survey of this literature, see La Porta and Shleifer (2014) and Ulyssea (2020)

[6]Please see Restuccia and Rogerson (2017) and Hopenhayn (2014) for the overwiew of misallocation literature.

[7]Access to the census data is restricted and requires clearance from INEGI. Data for this paper were processed at INEGI's data lab in Mexico City.

[8]Rural areas have less than 2,500 residents but INEGI does include rural areas in the census when they are the capital of the municipality.

data collection is not restricted to registered (or formal) businesses. Establishments are included regardless of their size (even establishments without employees), as long as the economic activity takes place within delimited fixed building structures.[9]

The census is administered every five years. We exploit the five waves between 1998 and 2018. We consider only establishments in manufacturing, wholesale and retail, and non-financial services.[10] In these three broad sectors, the number of establishments in the census increased from 2.5 million in 1998 to 4.3 million in 2018. In 2018, 14% of establishments were in manufacturing, 52% were in wholesale and retail, and the remaining 34% were in (other) non-financial services. Average establishment size in the census has hovered at about 4.5 employees (about 7.5 excluding establishments with no employees).[11] 96% of establishments employ 10 or fewer employees (a fraction that has not changed since 1998) and accounted for 44% of employment in 2018; less than 0.5% of establishments employ more than 100 workers but in 2018 accounted for 36% of employment (Table 2 below and Table B.3 in the appendix).

2.1 Formal and informal plants

The 2018 questionnaire introduced two new questions where the respondent is asked whether the establishment is registered with the central tax authority SAT (*Servicio de Administración Tributaria*) and with the social security administration IMSS (*Instituto Mexicano del Seguro Social*). SAT collects corporate and personal income taxes (similar to the IRS in the US), value added taxes, and federal excise taxes. IMSS, on the other hand, collects social security contributions, which employers are required both to contribute to and to withhold on behalf of their employees.[12] The two questions on registration with SAT and IMSS had yes-or-no answers, with the possibility of a refusal to respond.[13]

---

[9] For example, businesses located inside residential properties are included, but street vendors are not.

[10] To be precise, our analysis excludes agriculture (11 in the NAICS classification system); mining (21); utilities (22); construction (23); finance (52); real estate (53); management services and administrative support (55-56); education (61); healthcare and social assistance (62); postal services (491); religious and civic organizations (813); private households (814); and employment services (5613).

[11] We measure employment adding white collar and blue collar employees, unpaid workers, third-party employees (who don't have an official employment relationship to the establishment), and contractors.

[12] IMSS also delivers some social services. The share of employers' social security contributions amount to a tax of about 24% of profits (World Bank Group Doing Business 2020).

[13] In Spanish, the questions read *¿Este establecimiento cuenta con registro ante el Servicio de Administración Tributaria (SAT)?* and *¿Cuenta con registro patronal ante el Instituto Mexicano del Seguro Social (IMSS)?* Two follow-up questions asked the respondent to provide the identification numbers of the establishment with each tax authority. These four questions were not included in 1998-2013 waves.

A third of establishments in the 2018 cross-section report registration with SAT and only 0.23% have missing values in this question. In contrast, only 6.6% of plants report registration with IMSS, but the rate of non-response is 11.5% (see Table 1 below). As we show below, rates of non-response are correlated with the size of the plant—smaller establishments are less likely to report registration with tax authorities.

Table 1: Sorting of establishments into formal and informal and share of plants and employment in each category. 2018.

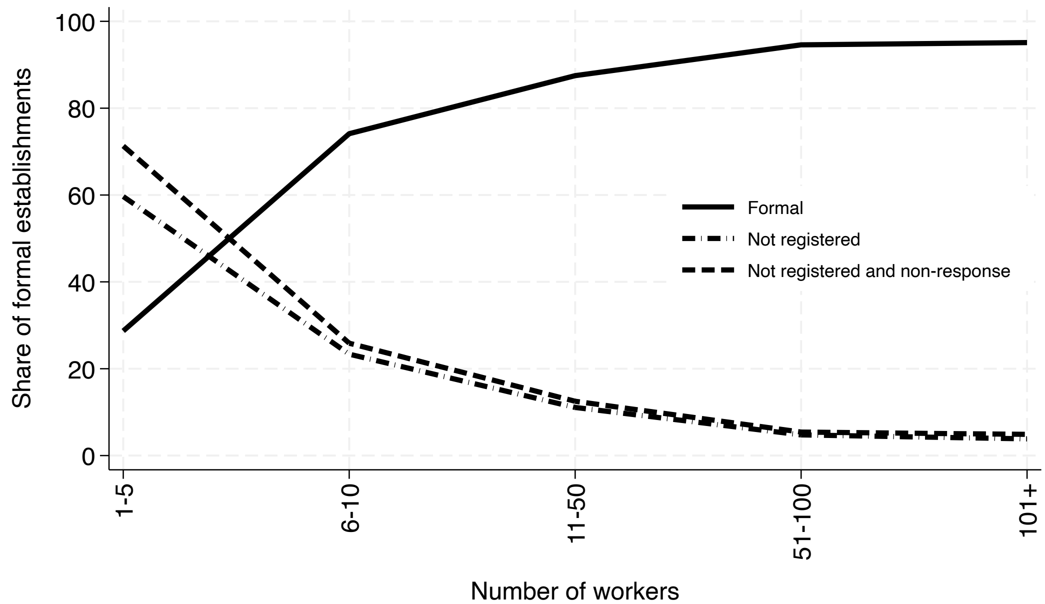
SAT	IMSS	Formal / informal	Share of plants	Share of workers
Yes	Yes	Formal	6.48	40.14
No	Yes	Formal	0.11	0.33
Yes	No	Formal	26.24	30.24
Yes	Non-response	Formal	0.76	0.33
Non-response	Yes	Formal	0.00	0.00
No	No	Informal	55.71	25.03
Non-response	Non-response	Informal	0.23	0.54
No	Non-response	Informal	10.48	3.39
Non-response	No	Informal	0.01	0.00

Notes: Total in each column may not add up to 100 due to rounding error.

In our analysis, establishments registered with SAT and/or IMSS (as reported by the respondent) are formal; establishments that are not formal are informal (unregistered) as we list in Table 1.[14] Non-responses are thus included in the informal category. Figure 1 shows, for each size bin, the fraction of registered or formal establishments in the 2018 cross-section. The share of formal establishments according to our definition markedly increases with size. The figure also shows that the share of plants not registered with either SAT or IMSS and the rate of non-response to the two questions is higher among small plants. Under our definition, 34% of establishments are formal, accounting for 71% of employment in the census.

[14] In comparison to the definition developed by Busso et al. (2012) and Levy (2018), which focuses on the intensive margin of informality (defined by Ulyssea (2018)), our analysis looks at informality on the extensive margin.

Figure 1: Share of formal and informal establishments in each size category. 2018.



Notes: Formal establishments are registered with SAT, IMSS, or both, as self-reported by the respondent. Establishments that are not formal are informal.

Table 2 shows the size distribution and the employment size distribution of formal and informal establishments in the 2018 cross-section. Micro-establishments (with 1-10 workers) amount to 90% of establishments and account for 26% of employment among formal plants. Among informal plants, micro-establishments amount to 99% of plants and account for 90% of employment. Large establishments (with 101 or more workers) account for almost 50% of employment among formal plants but only 4% of employment among informal establishments. Informal establishment size averages 2.03, while average formal establishment size is 10.1 workers.

Table 2: The size distribution of formal and informal plants in the Mexican census. 2018.

Size	**Total**		**Formal plants**		**Informal plants**	
	Share of plants	Share of workers	Share of plants	Share of workers	Share of plants	Share of workers
1-10	95.89%	43.70%	89.10%	26.00%	99.32%	89.22%
11-20	2.06%	6.22%	5.28%	7.57%	0.44%	2.98%
21-50	1.19%	7.84%	3.21%	9.99%	0.18%	2.65%
51-100	0.39%	5.81%	1.10%	7.75%	0.03%	1.09%
101+	0.46%	36.42%	1.31%	48.69%	0.03%	4.06%
Total	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Notes: Formal establishments are registered with SAT, IMSS, or both, as self-reported by the respondent. Establishments that are not formal are informal.

2.2 The life cycle of formal and informal plants

We estimate size at birth as the average size of formal and informal plants ages 0-4 in the 2018 cross-section: 6.1 workers among formal establishments and 1.9 among informal plants (2.9 for the total average). To estimate growth rates, we exploit the panel dimension of the data. INEGI has developed identifiers that track establishments across the 2008, 2013, and 2018 waves of the census. Busso et al. (2018) developed synthetic identifiers to follow establishments from 1998 through 2008, which we leverage in our analysis.[15] Busso et al. (2019) first exploited the synthetic identifiers to try to compute the life cycle of formal and informal plants in Mexico, but they focused on informality on the intensive margin while our definition looks at the extensive margin. Moreover, while they restrict their analysis only to the balanced panel (the panel of surviving plants between 1998 and 2018), our estimates of the average growth rates leverage micro data from entering and exiting establishments as well as in Eslava et al. (2022). Our analysis also corrects for differences across sectors.

We first sort establishments in the 2018 cross-section into formal and informal using the self-reported registration with SAT and IMSS as described above. We then treat establishments that were informal (formal) in 2018 as informal (formal) every period back through 1998. That is, we do not estimate separate profiles for plants that transition in and out of informality during their life cycle and for plants that never changed their

[15]Busso et al. (2018) use the location, the name, and the 6-digit sector to find establishments in two consecutive waves of the census.

status since we are not able to observe in the data transitions across informality status using our definition (the questions on registration with SAT and IMSS were not available in previous waves of the census).[16] Still, in our quantitative exercises, we do allow for establishments to change their informality status during their life cycle to assess the potential effect on their growth rates.

We sort formal and informal establishments in every period into age bins (0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, and 35-39). Next, we restrict the data to those establishments that were ages 0-4 in any cross-section between 1998 and 2013. That is, we restrict the analysis to establishments in the 2018 cross-section (when we observe their formality status) that started operations between 1998 and 2013 (3,006,254 unique establishments). For each establishment, we follow Eslava et al. (2022) and divide their size in every period by their size at age 0-4. We then regress these cumulative growth rates on the (double) interaction between age (in bins) and a binary indicator for whether the plant is formal or informal, 2-digit sector and cohort fixed effects, and another indicator for whether the year is before or after 2008, to control for whether the identifier that links establishments across waves is synthetic (as generated by Busso et al. (2018)) or developed by INEGI.

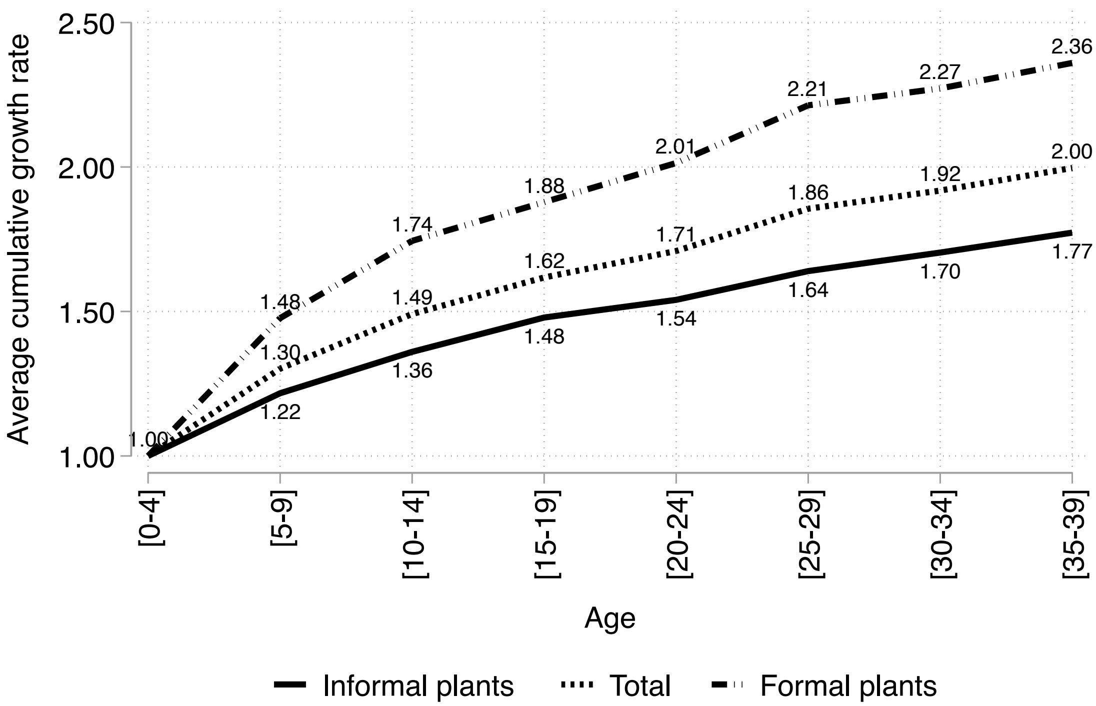
The exercise just described only allows us to obtain average growth rates through ages 20-24 since we only observe an establishment for 5 periods at most. We therefore repeat these steps for different starting ages. For example, to obtain the average growth rate through ages 30-34, we restrict the data to establishments in the 2018 cross-section that were 10-14 at any point between 1998 and 2013. Our life cycles, which we show in Figure 2, are the average of the different estimates for the cumulative growth rates from the separate regressions using different starting ages. More details on our methodology, together with the full set of results from least squares, are offered in the appendix.

Our estimates show that formal establishments more than double their size during their life cycle. In Figure 2, the size of formal establishments at ages 35-39 is on average 2.4 times their size at birth. In contrast, informal plants increase their size only by 77%. These estimated growth rates do not differ significantly when we focus only

[16] One potential proxy are the transitions across status of informality on the intensive margin. Busso et al. (2019) estimate that almost 90% of informal establishments in 1998 were still informal in 2013. Formality on the extensive margin, however, is more likely than informality on the intensive margin. While 34% of establishments is registered with SAT and/or IMSS, less than 10% of the 2018 cross-section employ salaried employees and fully contribute to their social security (the definition of formality on the intensive margin first coined by Busso et al. (2012)). Moreover, only 27% of formal establishments according to our definition are also formal on the intensive margin (Table B.4 in the appendix).

manufacturing or services (see B.2 in the appendix). Formal plants then not only start their life cycle with more workers relative to informal plants (three times more), but also grow at a markedly faster rate.

Figure 2: Cumulative growth rates in the life cycle of formal and informal plants in the Mexican census.



Notes: Formal establishments are registered with SAT, IMSS, or both, as self-reported by the respondent. Establishments that are not formal are informal. These cumulative growth rates are (rescaled) marginal effects from least squares regressions that also control for fixed effects for 2-digit sector and cohort, and a dummy for whether the panel identifier is synthetic (before or after 2008).

In Figure B.3 in the appendix, we compute growth rates for formal and informal plants according to our definition but exploiting the 2018 cross-section only, as in Hsieh and Klenow (2014). Interestingly, the cross-sectional estimates seem to bias downward the growth of informal establishments (40% in the cross-section vs. 77% exploiting the panel) but overestimate growth among formal plants (170% in the cross-section vs. 140% in our estimations). We show in Figures B.4 and B.5 the life cycles for formal and informal plants using different definitions of informality. Figure B.4 looks at a strict definition that does not sort non-responses into the informal category: plants registered with both SAT and IMSS are formal, while plants not registered with SAT nor IMSS are informal. Figure B.5 follows the formality definition on the intensive margin of Busso et al. (2012) and Levy (2018). The estimated growth rates for informal plants do not vary across definitions (from 77% in our extensive margin definition, to 64% under a stricter classification, and 69% in the intensive margin definition), but estimated growth rates among formal establishments are significantly higher under these alternative definitions, from 136% in our extensive margin definition, to 228% under a stricter classification, and 206% in the intensive margin definition of Busso et al. (2012) and Levy (2018).

3. Model

We build on the model developed by GPV where entrepreneurs run bigger plants by investing on their entrepreneurial ability. Our main innovation is introducing a tax on output that some plants can escape from by operating informally due to incomplete enforcement.

3.1 Environment

In every period, a group of agents endowed with entrepreneurial ability z is born. The initial endowment of entrepreneurial ability is drawn from the (exogenous) probability density function $f(z)$ with cumulative distribution function $F(z)$ and support $[\underline{z}, \overline{z}]$. Agents live for J periods.

Each agent maximizes the expected present value of lifetime utility derived from consumption:

$$\sum_{j=1}^{J} \beta^{j-1} u(c_j) \tag{1}$$

where $\beta \in (0, 1)$ is the discount factor, c_j denotes the consumption of an agent at age j, and $u(c) = \log(c)$.

Agents have one unit of time which they supply inelastically until they reach their retirement age $J_R < J$. Along their life-cycle, agents choose to be workers or entrepreneurs. Furthermore, entrepreneurs choose whether to operate formal plants, i.e., complying with all taxes, or informal plants, i.e., escaping from taxation at each period until the retirement age.

Production of the single final good of the economy takes place in plants and it requires entrepreneurial ability, z, capital, k, and labor, n, as follows:

$$f(z,k,n) = Az^{1-\gamma}\left[k^{\alpha}n^{1-\alpha}\right]^{\gamma}, \tag{2}$$

where γ is the span-of-control parameter, with $\gamma \in (0,1)$, indicating decreasing returns to scale. The parameter A denotes the aggregate productivity term common across all entrepreneurs.

Entrepreneurs can improve their entrepreneurial skills by investing their income according to the following skill accumulation technology:

$$z_{j+1} = z_j + g(z_j, x_j) \tag{3}$$

where $g(z_j, x_j) = z_j^{\theta_1} x_j^{\theta_2}$ and x_j denotes the income invested in entrepreneurial skill by an entrepreneur at age j. The parameters θ_1 and θ_2 control the importance of existing entrepreneurial skills and investment in entrepreneurial skill, respectively, where $0 < \theta_i < 1$ for $i \in \{1, 2\}$. Hence, the skill accumulation technology exhibits complementarity between the existent entrepreneurial skill and investments in skill ($g_{zx} > 0$) and diminishing returns to skill investments ($g_{xx} < 0$). Moreover, investment in entrepreneurial skill is an essential input for the skill accumulation technology ($g(z, 0) = 0$).

3.2 Taxation with incomplete enforcement

Entrepreneurs face a proportional output tax denoted by $\tau \in (0,1)$. Enforcement of this output tax, however, is incomplete for two reasons. First, at the beginning of their life cycle, agents are assigned to a tax official only with some exogenous probability $\bar{\phi} \in [0,1]$. If an entrepreneur faces a tax official, the output tax is fully enforced. In other words, agents who meet a tax official can only operate formally when they choose to become entrepreneurs.[17] These agents therefore have to decide only between becoming workers or formal entrepreneurs. This component of the enforcement technology does not affect the choice of capital or labor for entrepreneurs directly but does affect occupational choices.

If an agent does not face a tax official at the beginning of the life cycle, he or she can escape from taxation by operating informally (fully avoiding the output tax). Operating informally, though, carries the risk of detection and subsequent punishment (even if

[17] As we discuss in the calibration section, this probability of meeting with an official, inspired from Tamkoç (2024a), results in a positive fraction of small formal plants operating in equilibrium.

the individual does not face a tax official when they were born). The enforcement of taxation, in turn, follows a size-dependent technology:

$$p(k) = \begin{cases} 0 & \text{if} \quad k \leq b \\ 1 & \text{if} \quad \text{otherwise} \end{cases} \tag{4}$$

Informal entrepreneurs then can completely avoid paying taxes as long as they keep their capital equal or less than the threshold b. Note that no informal entrepreneur will then optimally choose to exceed this threshold as long as the fine if caught evading is strictly larger than zero. If the entrepreneur is informal, she will be get caught by the tax authorities when producing with more than b capital. In that case, the informal entrepreneur will be required to pay a fine in addition to the output tax, which would lead to lower profits compared to producing at $k = b$ or operating in the formal sector.

3.3 Problem of Entrepreneurs and Workers

In this section we describe the decisions of entrepreneurs and workers in a stationary environment where factor prices W and R are constant. We assume that both entrepreneurs and workers can lend and borrow assets, a, at the risk-free rate $r = R - \delta$.

A formal entrepreneur with ability z is subject to the output tax and chooses capital and labor to maximize her profit $\Pi^F(z)$:

$$\Pi^F(z) = \max_{k,n} (1 - \tau) A z^{1-\gamma} \left[k^\alpha n^{1-\alpha}\right]^\gamma - wn - Rk \tag{5}$$

Similarly, an informal entrepreneur with ability z chooses capital and labor to maximize her profit, $\Pi^I(z)$. Unlike formal entrepreneurs, informal entrepreneurs don't pay taxes but they are subject to a probability of getting caught:

$$\Pi^I(z) = \max_{k,n} (1 - p(k)) A z^{1-\gamma} \left[k^\alpha n^{1-\alpha}\right]^\gamma - wn - Rk \tag{6}$$

In a stationary equilibrium, if an agent chooses to become a worker at the beginning of her life-cycle, she will remain a worker until retirement, as she will not be accumulating entrepreneurial ability. If an agent chooses to become a formal entrepreneur at the beginning of her life-cycle, she will remain a formal entrepreneur until retirement since entrepreneurial abilities don't depreciate and the skill-investment technology doesn't feature stochastic returns. In contrast, agents who choose to become informal entrepreneurs at the beginning of their life-cycle may transition into the formal sector as they keep accumulating their abilities. We assume that this transition from informal to formal entrepreneurship is costless.

The income of an entrepreneur is the profit from running plants plus the asset income until she retires. After retirement, her income consists of only asset income. Therefore, an entrepreneur with initial ability z who runs informal plants until the working age $\hat{j}$ and transitions into running formal plants after age $\hat{j}$ chooses consumption, savings, investment on her entrepreneurial ability to maximize the present value of life-time utility:

$$\begin{aligned} V^{J_R-\hat{j},F}(z) = \max_{c_j,x_j,a_{j+1}} \quad & \sum_{j=1}^{J} \beta^{j-1} u(c_j) \\ \text{s.t.} \quad & c_j + x_j + a_{j+1} = \Pi^I(z_j) + (1+r)a_j, \quad \forall\, 1 \leq j < \hat{j} - 1 \\ & c_j + x_j + a_{j+1} = \Pi^F(z_j) + (1+r)a_j, \quad \forall\, \hat{j} \leq j \leq J_R - 1 \\ & c_j + a_{j+1} = (1+r)a_j \quad \forall j \geq J_R \\ & z_{j+1} = z_j + g(z_j, x_j), \quad \forall\, 1 \leq j < J_R - 1 \end{aligned} \tag{7}$$

Therefore, an entrepreneur with ability z chooses also at what age she transitions into the formal sectors: $V^E(z) = \max\{V^{0,F}(z), V^{1,F}(z), ..., V^{J_R-1,F}(z)\}$ such that $V^{0,F}(z)$ denotes the value of being a formal entrepreneur throughout the working life-time and $V^{J_R-1}(z)$ denotes the value of being an informal entrepreneur until retirement.

The problem of a worker is choosing consumption and savings at each age to maximize the present value of life-time utility.

$$\begin{aligned} W = \max_{c_j,a_{j+1}} \quad & \sum_{j=1}^{J} \beta^{j-1} u(c_j) \\ \text{s.t.} \quad & c_j + a_{j+1} = w + (1+r)a_j, \quad \forall\, 1 \leq j \leq J_R - 1 \\ & c_j + a_{j+1} = (1+r)a_j \quad \forall j \geq J_R \end{aligned} \tag{8}$$

Workers supply their time inelastically and receive a market wage, irrespective of their initial entrepreneurial skills and of whether they work for a formal or an informal entrepreneur. Hence, workers' income includes wages and asset returns until retirement. After retirement, their only income comes from asset holdings similar to retired entrepreneurs.

3.4 Occupation Choices

Agents select the occupation that results in the maximum life-time utility. An agent with ability z chooses to become an entrepreneur if $V^E(z) \geq W$; otherwise, she becomes a worker. Furthermore, since an agent with ability z who is not assigned to the tax official has an option to operate informally, she further chooses whether to operate formally or informally at each age. Appendix C provides the first-order condition of entrepreneurs' factor input choices and the definition of the equilibrium.

3.5 Discussion of Equilibrium Properties

In this section, we discuss properties of the stationary equilibrium. First, the profits of a formal entrepreneur with ability z can be expressed as the function of parameters and factor prices by using optimal factor demands, Equations (C.1) and (C.2), as follows:

$$\Pi^F(z) = [(1-\tau)A(1-\gamma)^{1-\gamma}\gamma^{\gamma}]^{\frac{1}{1-\gamma}} \left(\frac{\alpha}{R}\right)^{\frac{\alpha\gamma}{1-\gamma}} \left(\frac{1-\alpha}{w}\right)^{\frac{(1-\alpha)\gamma}{1-\gamma}} z \tag{9}$$

In equilibrium, some informal entrepreneurs will choose a level of k strictly below b, but others will opt to adjust their capital level precisely to b, maintaining just enough capital to remain undetected. We refer to informal entrepreneurs operating with capital levels less than b as unconstrained entrepreneurs, an informal entrepreneurs operating exactly at b as constrained informal entrepreneurs.

Unconstrained informal entrepreneurs choose labor and capital (which is less than b) while the constrained informal entrepreneurs choose labor given that their capital is b. Using equations C.3, C.4, and C.5, the profit of informal entrepreneurs can be expressed as follows:

$$\Pi^I(z) = \begin{cases} [A(1-\gamma)^{1-\gamma}\gamma^{\gamma}]^{\frac{1}{1-\gamma}} \left(\frac{\alpha}{R}\right)^{\frac{\alpha\gamma}{1-\gamma}} \left(\frac{1-\alpha}{w}\right)^{\frac{(1-\alpha)\gamma}{1-\gamma}} z & \text{if} \quad k^I(z) \leq b \\ (1-(1-\alpha)\gamma) \left[Ab^{\alpha\gamma} \left(\frac{(1-\alpha)\gamma}{w}\right)^{\gamma(1-\alpha)}\right]^{\frac{1}{1-(1-\alpha)\gamma}} z^{\frac{1-\gamma}{1-\gamma(1-\alpha)}} - bR & \text{if} \quad k^I(z) = b \end{cases} \tag{10}$$

Notice that the profit function of unconstrained informal entrepreneurs (i.e., when $k^I < b$) is linear in z, while the profit function of constrained informal entrepreneurs is concave in z as $\frac{1-\gamma}{1-\gamma(1-\alpha)} < 1$.

The inter-temporal decision between today's consumption and the next period's asset can be characterized by the Euler equation:

$$\frac{1}{c_j} = \beta(1+r)\frac{1}{c_{j+1}} \qquad \forall 1 \leq j \leq J-1 \tag{11}$$

which indicates that agents are indifferent between consuming at age j and saving to consume at age $j+1$ in equilibrium.

Entrepreneurs also decide how much to invest on their entrepreneurial skills. This decision can be found from the no-arbitrage condition for asset and investment in entrepreneurial ability as follows:

$$(1+r) = g_x(z_j, x_j)\Pi_z^i(z_{j+1}) + \frac{g_x(z_j, x_j)}{g_x(z_{j+1}, x_{j+1})}[1 + g_z(z_{j+1}, x_{j+1})], \quad \forall 1 \leq j < J_R - 2 \tag{12}$$

and

$$(1+r) = g_x(z_j, x_j)\Pi_z^i(z_{j+1}), \quad \text{for} j = J_R - 2 \tag{13}$$

The left-hand side of Equation (12) represents the opportunity cost of investing in entrepreneurial skills. In other words, an entrepreneur could earn $(1+r)$ units of the consumption good by investing one unit in assets rather than investing one unit in entrepreneurial skills. The right-hand side of the equation is the net marginal benefit of skill investment. The first term represents the benefit deduced from the next period profit generated by additional entrepreneurial ability. An extra unit of skill investment in the current period is transformed into next period entrepreneurial skill with $g_x(z_j, x_j)$ and $\Pi_z^i(z_{j+1})$ transforms the additional entrepreneurial ability to profit. Also, an extra unit of investment in entrepreneurial ability in a given age affects skill accumulation decisions in all subsequent periods. The second term in the right-hand side of Equation (12) represents this effect. This term disappears in the last working age period of entrepreneurs as shown in Equation (13).

Occupational choices can be determined by a set of threshold abilities. First, for agents assigned to a tax official, the occupational choice is characterized by a single threshold ability, $z_{w,e}$. Agents with abilities below this threshold become workers, while those above the threshold become formal entrepreneurs. For agents not assigned to a tax official, there are J_R threshold ability levels due to the additional option of operating informally and the possibility of transitioning to formal entrepreneurship over their life cycle.[18] Let $[z_{j-1,F}, z_{j,F}]$, for $1 \leq j \leq J_R - 1$, denote the set of agents who choose

[18]Without loss of generality, we refer to formal and informal entrepreneurship, including the transition

to be informal entrepreneurs at the beginning of their life-cycle but transition into being a formal entrepreneur at the last $j-1$ periods of their working life, such that $[\underline{z}, z_{0,F}]$ is the set of workers who are not assigned to a tax official. For example, the set of agents with ability in $[z_{0,F}, z_{1,F}]$ always operate informally (never transition into formal entrepreneurship) while agents with abilities in $[z_{1,F}, z_{2,F}]$ operate informally until the last period of their working life. Lastly, entrepreneurs with abilities in $[z_{J_R,F}, \bar{z}]$ operate formally their entire working life. Hence there are at most J_R threshold levels that determines the occupation choice of agents who are not assigned to a tax official: $z_{0,F} \leq z_{1,F} \leq ... \leq z_{J_R-2,F} \leq z_{J_R-1,F}$. Notice that there may not be any agent who would like to transition into formality at the last periods of their working life as they don't have much time to invest on their entrepreneurial ability and grow. In these cases, the occupation choices of agents who are not assigned to a tax official are characterized by fewer threshold ability levels. In sum, there are at most $J_R + 1$ threshold ability levels that determine the occupational choices in equilibrium (one for agents who are assigned to a tax official and J_R for those who are not assigned to a tax official).

4. Calibration

We set the model period to 5 years, corresponding to the intervals used in our data analysis. In this framework, agents work for 40 years (8 model periods) and then spend 15 years in retirement (3 model periods). We set the following parameter values from Leal (2014): depreciation rate $\delta = 0.055$, discount factor $\beta = 0.945$, and tax rate $\tau = 0.25$. After calibrating the span-of-control parameter γ, we determine α such that the importance of capital in the production ($\alpha\gamma$) equals 0.33, the standard value in the development literature. We assume that entrepreneurial ability follows a log-normal distribution with parameters μ_z and σ_z. We normalize the mean parameter $\mu_z = 0$ and productivity parameter $A = 1$.

These parametric assumptions leave six parameters to be determined: $\gamma, \sigma_z, \theta_1, \theta_2, b$, and $\overline{\phi}$. We jointly calibrate these parameters to match key features of the establishment distribution in the Mexican census data, with particular attention to the low growth rate of informal plants. Our calibration targets the following moments: (i) the overall mean establishment size; (ii) the mean size of informal establishments; (iii) the fraction of establishments with fewer than 10 workers; (iv) the employment share of establishments with more than 50 workers; (v) the life-cycle growth rate of formal establishments (from

out of informal entrepreneurship, as different occupations.

ages 0-4 to 35-39); and (vi) the life-cycle growth rate of informal establishments (from ages 0-4 to 35-39).

To align our model moments with data moments, we make several adjustments. Our definition of informality is based on establishments' registration status with SAT and IMSS in the last year of the census. Therefore, when calculating moments related to the plant-size distribution in the model, we classify retiring entrepreneurs according to their operational status: those retiring as informal entrepreneurs are counted as informal plants, and similarly for formal plants. The mean size of informal plants is defined as follows:

$$\text{Informal Mean Size} = \frac{\sum_{j=1}^{8} \int_{z_{0,F}}^{z_{1,F}} n^{0,F}(z,j) f(z) dz}{\sum_{j=1}^{8} \int_{z_{0,F}}^{z_{1,F}} dz} \tag{14}$$

where $n^{0,F}(z,j)$ represents the labor demand of an informal entrepreneur with initial ability z at age j. In Equation (14), the numerator represents the total labor demand by informal entrepreneurs across all eight generations, while the denominator represents the fraction of informal entrepreneurs in equilibrium.

To maintain consistency with the data, we focus on entrepreneurs who remain in their initial sector (formal or informal) when calculating growth rates. The growth rate of informal plants from age 0-4 to 35-39 is defined as:

$$g_{1-8}^{I} = \frac{\int_{z_{0,F}}^{z_{1,F}} (\frac{n^{0,F}(z,8)}{n^{0,F}(z,1)} - 1) f(z) dz}{F(z_{1,F}) - F(z_{0,F})} \tag{15}$$

Hence, g^I represents the average growth rate of informal establishments, analogous to our empirical measure.[19] The growth rate for formal entrepreneurs follows a similar definition.

Table 3 shows the calibrated parameters whereas the targeted moments and their calibrated counterparts are shown in Table 4. The model matches the overall mean size and smaller informal mean size very successfully. It also matches the growth rates of informal and formal plants closely. Moreover, the model doesn't have any problem capturing the large concentration of employment in bigger plants and the large fraction

[19] The growth rate measured in the cross-sectional data differs from the average growth rate of individual establishments, both in the data and the model. The growth rate in the cross-section would be defined as $\frac{\int_{z_{0,F}}^{z_{1,F}} n^{0,F}(z,8) f(z) dz}{\int_{z_{0,F}}^{z_{1,F}} n^{0,F}(z,1) f(z) dz} - 1$.

of plants with less than 10 workers.

Table 3: Parameters Values

Parameter	Value
Technology level (A)	1
Depreciation rate (δ)	0.055
Discount factor (β)	0.945
Output tax (τ)	0.25
Importance of capital (α)	0.56
Mean of log entrepreneurial ability (μ_z)	0
Span-of-control (γ)	0.59
Dispersion in entrepreneurial ability (σ_z)	1.97
Return on previous productivity (θ_1)	0.58
Return on productivity investment (θ_2)	0.56
Enforcement threshold (b)	1.91
Probability of meeting official at birth ($\overline{\phi}$)	0.31

Note: This table summarizes all model parameters. First six parameters are set exogenously and the last six parameters are calibrated jointly.

Table 4: Model Performance

Moment	Data	Model
Overall mean size	4.77	4.77
Informal mean size	2.03	2.03
Formal growth (0-4 to 35-39) (%)	136.03	136.03
Informal growth (0-4 to 35-39) (%)	77.28	77.28
Fraction of plants <10 workers (%)	95.89	95.51
Emp. share of plants with +50 workers (%)	42.23	42.08

Source: INEGI
Note: This table compares moments of the data with that of the benchmark calibration given the parameters values in Table 3.

Next, we evaluate the benchmark economy's performance by examining non-targeted moments and its detailed features. We begin by comparing the plant size distributions in the model and the data. Figures 3a and 3b contrast the size distribution in the Mexican census with that in the calibrated economy. The benchmark model successfully replicates two key features of the Mexican census data: the large share of micro-plants (those with fewer than 10 workers) and the high concentration of employment in both micro-plants and large plants (those with more than 100 workers).

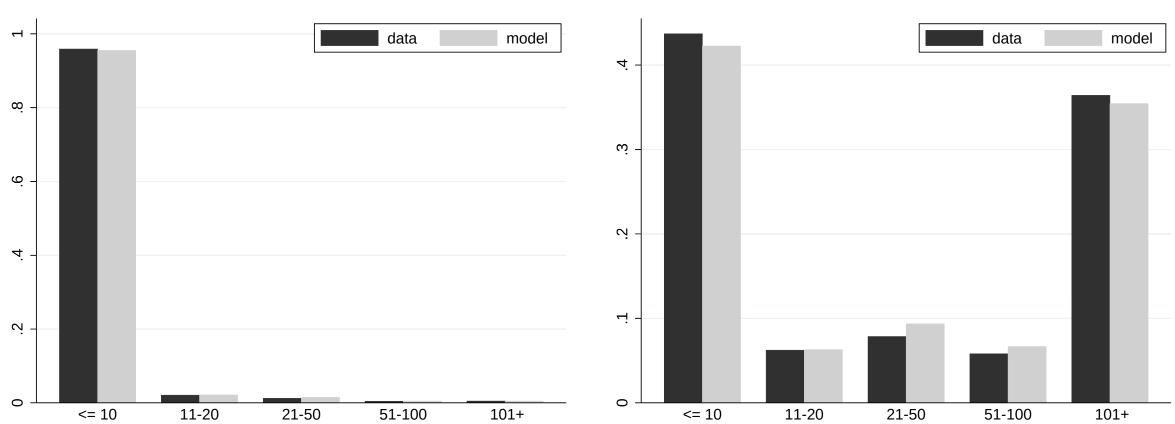
The model also approximates the formal plant size distribution. Figures 4a and 4b show that the model closely matches the fraction of formal plants among smaller plants (those with fewer than 10 workers). While the model accurately reflects the majority of plants operating in the informal sector, as observed in the data, it slightly overestimates the number of informal plants. In the model, the fraction of informal plants is around 77%, compared to 66% in the data. Similarly, the mean size of formal plants in the model is 14.1 workers, while in the data, it is 10.1. To sum up, only small fraction of plants are formal and they are bigger in terms of employment compared to informal plants. Formal plants account for 67.1% of all employment and 72.5% of aggregate output in the benchmark economy.

Before comparing the non-targeted life-cycle profiles in the model and the data, we first discuss the mean sizes at the initial period. In the data, informal plants employ an average of 1.9 workers at birth, whereas in the model, the mean size of informal plants at birth is 1.5 workers. The model also closely matches the overall mean size at birth. In Mexico, the overall mean size at birth is 2.1 workers, compared to 2.9 workers in the model.

Figure 3: Plant size distribution in the benchmark economy

(a) Fraction of Plants

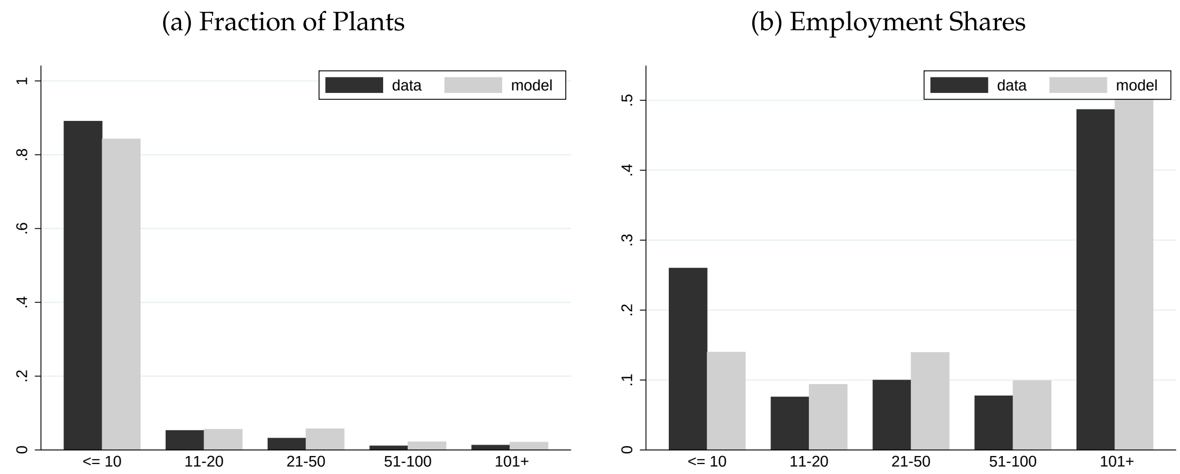
(b) Employment Shares



Source: INEGI
Notes: This figure compares the overall plant size distribution in the benchmark economy with the data for different size categories.

Next, we compare the life-cycle profiles of formal and informal plants in the benchmark equilibrium with those observed in the data. Figure 5 shows that the model is able to replicate the life cycle of the average plant in Mexico. It also closely matches the life cycle of informal plants observed in the data. However, for formal plants, the

Figure 4: Formal plant size distribution in the benchmark economy



Source: INEGI
Notes: This figure compares the formal plant size distribution in the benchmark economy with the data for different size categories.

model predicts a lower growth profile compared to the data, resulting in an overall growth rate in the model that is slightly lower than in the data. The growth rates in the cross-section of plants shows similarities with the data. In the model, the growth rates of formal plants in the cross-section are higher than their average growth rates, while the growth rates of informal plants in the cross-section are lower than their average growth rates, as in the data. Informal plants in the cross-section grow by 61.3% over their life cycle (from the 0-4 to the 35-39 age group) in the model, compared to a 583.5% growth rate for formal plants in the cross-section

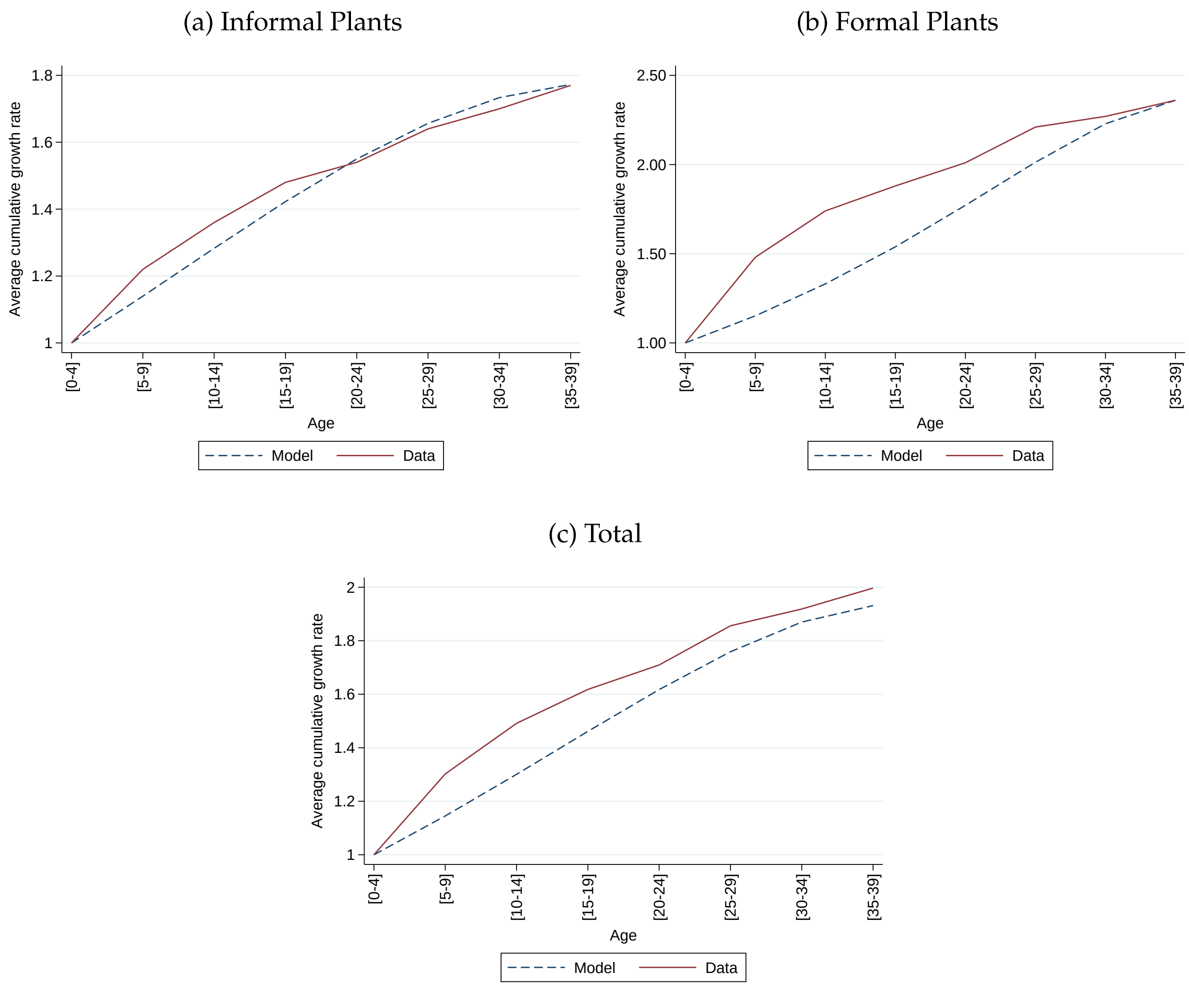
Finally, we discuss informal plants transitioning into formality over their life cycle before presenting the quantitative exercises. In the benchmark economy, no informal plants transition into formality during the last or the last two periods of their life cycle. Only 1.9% of plants begin their life as informal and switch to formality during their life cycle. Therefore, the occupation choices of agents are characterized by eight threshold ability levels in the benchmark economy.

5. Quantitative Findings

The development literature mainly focuses on the role of reducing the burden of being formal and increasing enforcement on the prevalence of informality.[20] In this section,

[20] Please see Kuehn (2014), Orsi et al. (2014), Leal (2014), Ulyssea (2018), Johnson et al. (1998), Fajnzylber et al. (2011), De Mel et al. (2013), Bruhn and McKenzie (2014), De Andrade et al. (2016) and Rocha et al.

Figure 5: Life-cycle of informal and formal plants in the benchmark



Source: INEGI

we first present the equilibrium effects of changes in the tax rate, τ and changes in the enforcement levels, b and $\bar{\phi}$. Then we discuss the implications of revenue-neutral enforcement exercises where tax rates are adjusted to keep tax revenues constant as enforcement levels change.

In this economy, the only burden of being formal is the output tax, τ. Table 5 shows how the model behaves under different values of the tax rate. As the output tax increases, more plants choose to operate informally to escape from high taxation. Doubling the output tax increases the share of informal plants by 13.1 percentage points, and decreases mean size and aggregate output by 36.1% and 37.6%, respectively. The

(2018) among many others.

overall plant growth rate in the economy rises as more productive formal plants in the benchmark decide to operate informally. On the other hand, removing all output taxes more than doubles the mean size and average entrepreneurial quality compared to the benchmark. When there is no output tax in the economy, aggregate output increases by 46.5% and the life-cycle growth rate increases by 85%.

As τ decreases, the share of informal plants transitioning into formality over their life-cycle increases. For example, when the output tax is reduced by half (from 25 to 12.5%), the fraction of transitioning informal plants increases to 5.8% compared to 1.9% in the benchmark. On the other hand, a higher tax rate generates stronger incentives to avoid it, which can happen in this economy by investing less in entrepreneurial skill to remain small. This results in a higher rate of informality, a lower average plant size, and a worse selection into entrepreneurship (lower average entrepreneurial ability).

Table 5: Effects of Output Tax

			Benchmark		
	$\tau = 0$	$\tau = 0.125$	$\tau = 0.25$	$\tau = 0.375$	$\tau = 0.50$
Mean size	9.83	6.85	4.77	3.56	3.05
Informal mean size		1.51	2.03	2.56	2.94
Frac. Of informal plants (%)	0	62.84	77.32	84.77	90.42
Aggregate output	146.49	122.37	100	78.83	62.38
Employment share of 50+ (%)	62.85	55.30	42.08	21.94	4.20
Growth rate, 0-4 to 35-39 (%)	172.28	109.07	93.12	102.08	119.37
Entrepreneurial Quality	269.14	163.17	100	65.16	48.73
Tax revenue	0.00	73.39	100	81.81	39.23

Note: This table presents the effects of different output taxes, τ's, on the benchmark economy when holding other parameters constant. The aggregate output, the entrepreneurial quality and the tax revenue at the benchmark case are normalized to 100 for ease of exposition.

Table 6 shows how changes in the enforcement threshold b in this economy would affect plant sizes, aggregate output, selection into entrepreneurship (average entrepreneurial ability), and plant growth. When the output tax is fully enforced ($b = 0$), establishments have no incentive to remain small to avoid detection, and average size increases by 46.1% relative to the benchmark. Moreover, while the average growth rate remains almost the same, the average entrepreneurial quality increases by 30.2%. The effect of full enforcement on aggregate output is less than 1 percent.

Another way of increasing enforcement is through the probability of agents meeting with a tax official at birth, $\bar{\phi}$ in the model. Since agents who encounter a tax official do not have the option to escape taxation by operating informally, any increase in $\bar{\phi}$

Table 6: Role of enforcement by b

	$0 * b_{bench}$	$0.5 * b_{bench}$	Benchmark $b_{bench} = 1.91$	$1.5 * b_{bench}$	$2 * b_{bench}$
Mean size	6.97	4.62	4.77	4.90	5.03
Informal mean size		1.35	2.03	2.50	2.89
Frac. Of informal plants (%)	0	72.94	77.32	78.65	79.36
Aggregate output	100.84	99.49	100	100.60	100.77
Employment share of 50+ (%)	48.51	45.86	42.08	38.98	36.08
Growth rate, 0-4 to 35-39 (%)	92.15	68.97	93.12	113.15	127.57
Entrepreneurial Quality	130.20	97.45	100	102.05	103.72
Tax revenue	138.97	113.24	100	91.40	84.46

Note: This table presents the effects of different enforcement levels, b's, on the benchmark economy when holding other parameters constant. The aggregate output, the entrepreneurial quality and the tax revenue at the benchmark case are normalized to 100 for ease of exposition.

implies a stricter enforcement level. The case where all agents meet with a tax official, i.e., $\bar{\phi} = 1$, corresponds to full enforcement, which is also achieved by $b = 0$. Table 7 presents the quantitative role of $\bar{\phi}$. When no agents encounter a tax official, the fraction of informal plants increases, and as a result, the mean size, entrepreneurial quality, and overall growth rate decline.

Table 7: Role of enforcement by $\bar{\phi}$

	$0 * \bar{\phi}_{bench}$	$0.5 * \bar{\phi}_{bench}$	Benchmark $\bar{\phi}_{bench} = 0.30$	$1.5 * \bar{\phi}_{bench}$	$2 * \bar{\phi}_{bench}$
Mean size	4.18	4.45	4.77	5.16	5.54
Informal mean size	1.94	1.98	2.03	2.08	2.13
Frac. Of informal plants (%)	95.41	87.11	77.32	65.65	51.02
Aggregate output	99.64	99.87	100	100.42	100.66
Employment share of 50+ (%)	39.86	40.98	42.08	43.52	44.81
Growth rate, 0-4 to 35-39 (%)	89.68	91.59	93.12	94.12	94.27
Entrepreneurial Quality	91.61	95.54	100	105.69	110.89
Tax revenue	84.83	92.35	100	108.38	116.81

Note: This table presents the effects of different probabilities of being assigned to a tax official, $\bar{\phi}$'s, on the benchmark economy when holding other parameters constant. The aggregate output, the entrepreneurial quality and the tax revenue at the benchmark case are normalized to 100 for ease of exposition.

6. Revenue Neutral Better Enforcement

Table 8 shows the results of a policy experiment where we reduce the enforcement threshold b while we adjust the tax rate relative to the benchmark to keep the tax revenue constant. When there is a full enforcement, mean size increases 63.5% (from 4.77 to 7.8). Average entrepreneurial ability increases by 68.1%, and the average life cycle growth rate increases by 25.2%. As a result, aggregate output increases by 16.4%.

Table 8: Tax Revenue Neutral enforcement experiments

	$0 * b_{bench}$	$0.25 * b_{bench}$	$0.5 * b_{bench}$	$0.75 * b_{bench}$	Benchmark $b_{bench} = 1.91$
Mean size	7.80	6.31	5.62	5.29	4.77
Informal mean size		0.70	1.14	1.56	2.03
Frac. Of informal plants (%)	0.00	42.65	64.22	71.71	77.32
Aggregate output	116.35	114.15	110.72	107.04	100.00
Employment share of 50+ (%)	54.01	53.75	51.73	48.43	42.08
Growth rate, 0-4 to 35-39 (%)	116.60	93.93	81.43	84.62	93.12
Entrepreneurial Quality	168.13	141.00	125.45	115.62	100.00
τ	0.16	0.17	0.19	0.21	0.25

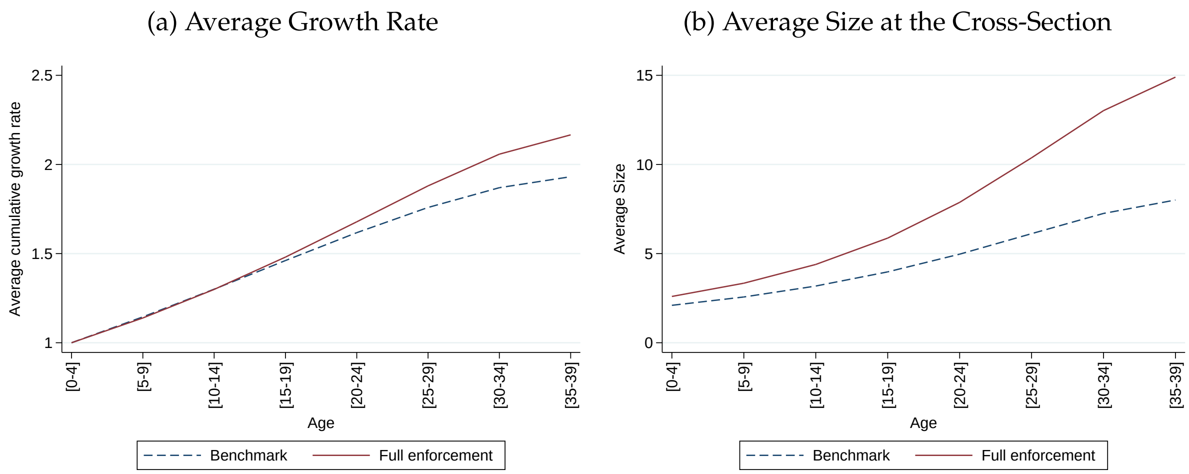
Note: This table presents the effects of stricter enforcement while keeping the tax revenue constant at the benchmark level by adjusting the output tax rate, τ . The aggregate output, the entrepreneurial quality and the tax revenue at the benchmark case are normalized to 100 for ease of exposition.

Figure 6 presents the effects of revenue-neutral full enforcement on average cumulative growth rates and average sizes across the cross-section of plants over their life cycle. When plants are younger (up to age 15-19), they don't benefit from full enforcement in terms of average cumulative growth rates as much as older plants. This is because the increase in input prices under full enforcement offsets some of the benefits of lower taxes when plants are younger and smaller. However, this doesn't mean their size remains unaffected by the policies. While the growth rates in the cross-section are overstated, we can use the average size at the cross-section to assess the policy's impact over the life cycle. Figure 6b shows that as a result of revenue-neutral full enforcement, plants of all ages increase their size, with older plants growing larger compared to younger plants, consistent with the average cumulative growth rate results.

7. Concluding Remarks

In this paper, we exploit data from Mexico's census plant-level data to quantify the output losses that arise from the marked differences in growth rates between formal

Figure 6: Effect of revenue-neutral full enforcement on the life-cycle



and informal plants. Unlike Hsieh and Klenow (2014), we are able to focus on a panel of registered and unregistered plants to estimate their life cycle employment profiles. We build on the model developed by Guner et al. (2018) where plants grow by investing on their productivity. Our innovation is introducing a tax on output that some plants can escape from by operating informally due to incomplete enforcement. Our policy experiments suggest that reducing the tax rate while improving the enforcement technology to keep the tax revenue constant would lead to output gains of 16.4% relative to the benchmark. It also implies increase in the average growth rate by 25.3%.

References

BARTELSMAN, E., J. HALTIWANGER, AND S. SCARPETTA (2013): "Cross-country differences in productivity: The role of allocation and selection," *American economic review*, 103, 305–334.

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

BUSSO, M., M. V. FAZIO, AND S. L. ALGAZI (2012): "(In) formal and (un) productive: the productivity costs of excessive informality in Mexico," Tech. rep., IDB working paper series.

BUSSO, M., O. FENTANES, AND S. LEVY (2018): "The longitudinal linkage of Mexico's economic census 1999-2014," *InterAmerican Development Bank*.

BUSSO, M., S. LEVY, AND J. TORRES (2019): "Establishment dynamics and the persistence of resource misallocation in Mexico: An analysis of longitudinal data for 1998-2013," *InterAmerican Development Bank*.

DE ANDRADE, G. H., M. BRUHN, AND D. MCKENZIE (2016): "A helping hand or the long arm of the law? Experimental evidence on what governments can do to formalize firms," *The World Bank Economic Review*, 30, 24–54.

DE MEL, S., D. MCKENZIE, AND C. WOODRUFF (2013): "The demand for, and consequences of, formalization among informal firms in Sri Lanka," *American Economic Journal: Applied Economics*, 5, 122–150.

DE SOTO, H. (1989): "The other path: The invisible revolution in the third world," .

ESLAVA, M., J. HALTIWANGER, AND Á. PINZÓN (2022): "Job Creation in Colombia Versus the USA:'Up-or-out Dynamics' Meet 'The Life Cycle of Plants'," *Economica*, 89, 511–539.

FAJNZYLBER, P., W. F. MALONEY, AND G. V. MONTES-ROJAS (2011): "Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program," *Journal of Development Economics*, 94, 262–276.

Garcia-Macia, D., C.-T. Hsieh, and P. J. Klenow (2019): "How destructive is innovation?" *Econometrica*, 87, 1507–1541.

Guner, N., A. Parkhomenko, and G. Ventura (2018): "Managers and productivity differences," *Review of Economic Dynamics*, 29, 256–282.

Hopenhayn, H. A. (2014): "Firms, misallocation, and aggregate productivity: A review," *Annu. Rev. Econ.*, 6, 735–770.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 124, 1403–1448.

——— (2014): "The life cycle of plants in India and Mexico," *The Quarterly Journal of Economics*, 129, 1035–1084.

Johnson, S., D. Kaufmann, and P. Zoido-Lobaton (1998): "Regulatory discretion and the unofficial economy," *The American economic review*, 88, 387–392.

Kuehn, Z. (2014): "Tax rates, governance, and the informal economy in high-income countries," *Economic Inquiry*, 52, 405–430.

La Porta, R. and A. Shleifer (2008): "The unofficial economy and economic development," Tech. rep., National Bureau of Economic Research.

——— (2014): "Informality and development," *Journal of economic perspectives*, 28, 109–126.

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

Levy, S. (2018): *Under-rewarded efforts: The elusive quest for prosperity in Mexico*, Inter-American Development Bank.

Loayza, N. V. (2016): "Informality in the Process of Development and Growth," *The World Economy*, 39, 1856–1916.

López, J. J. and J. Torres (2020): "Size-dependent policies, talent misallocation, and the return to skill," *Review of Economic Dynamics*, 38, 59–93.

Lucas, R. E. (1978): "On the size distribution of business firms," *The Bell Journal of Economics*, 508–523.

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

MEGHIR, C., R. NARITA, AND J.-M. ROBIN (2015): "Wages and informality in developing countries," *American Economic Review*, 105, 1509–1546.

ORSI, R., D. RAGGI, AND F. TURINO (2014): "Size, trend, and policy implications of the underground economy," *Review of Economic Dynamics*, 17, 417–436.

RESTUCCIA, D. AND R. ROGERSON (2008): "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic dynamics*, 11, 707–720.

——— (2017): "The causes and costs of misallocation," *Journal of Economic Perspectives*, 31, 151–74.

ROCHA, R., G. ULYSSEA, AND L. RACHTER (2018): "Do lower taxes reduce informality? Evidence from Brazil," *Journal of Development Economics*, 134, 28–49.

TAMKOÇ, M. N. (2024a): "Bribery, plant size and size dependent distortions," *Journal of Development Economics*, 103361.

——— (2024b): "Fading Away Informality by Development," Tech. rep., The World Bank.

ULYSSEA, G. (2018): "Firms, informality, and development: Theory and evidence from Brazil," *American Economic Review*, 108, 2015–2047.

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

Appendix

A. Estimating growth in the life cycle of formal and informal plants

To estimate average growth rates for formal and informal plants, we proceed as follows:

1. We sort establishments alive in 2018 into formal and informal, and apply this status back in the life cycle of each establishment (thus ignoring potential transitions into or out of informality).

2. In every period (census wave), we sort formal and informal establishments into 8 age bins: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, and 35-39. We then restrict the data to those that were in age bin 1 in any of the four waves prior to 2018. That is, we restrict the analysis in this first stage to establishments both (*i*) alive in 2018 (when we observe their formality status), and (*ii*) that were 0-4 in any period between 1998 and 2013 (prior to 2018). This results in 3,006,254 establishments that we can track from birth through 2018.

3. We compute cumulative growth rates for each establishment by dividing their size in every period by the size at birth (when they were in the first age bin). For example, for an establishment ages 0-4 in 2003, the cumulative growth rate in 2003 will be 1; the cumulative growth rate between ages 0-4 (age bin 1) and 5-9 (age bin 2) will be the size in 2008 divided by the size in 2003; the cumulative growth rate between ages 0-4 (age bin 1) and 10-14 (age bin 3) will be the size in 2013 divided by the size in 2003; and the cumulative growth rate between ages 0-4 (age bin 1) and 15-19 (age bin 4) will be the size in 2018 relative to the size in 2003.

4. We regress the cumulative growth rates on 2-digit sector and cohort fixed effects, a binary indicator for whether the cumulative rate corresponds to the period before or after 2008 (since before 2008, we only observe the synthetic identifiers generated by Busso et al. (2018)), and the (double) interaction of the indicator for whether the plant is formal or informal with the fixed effects for the age bins. We use the package *reghdfe* in STATA to absorb the sector and cohort fixed effects. The first column of Table A.1 below shows the results.

5. To obtain the life cycle of formal and informal plants and the average (or combined) life cycle, we take the derivative of the estimated equation using the STATA

command *margins*. The first column of Table A.2 shows the results. We rescale these resulting effects by dividing by the value at ages [0-4].

By construction, this exercise will only allow us to obtain the cumulative growth rates through ages 20-24 (since we only observe an establishment for 5 periods at most). We therefore repeat steps 1-5 above for different starting ages:

- We restrict the data to establishments alive in 2018 that were in age bin 2 (5-9) at any point between 1998 and 2013. This way, we can compute growth rates from ages 5-9 through 25-29.
- To obtain the life cycle through ages 30-34, we restrict the data to establishments alive in 2018 that were in age bin 3 (10-14) at any point between 1998 and 2013.
- To obtain the life cycle through ages 35-39, we restrict the data to establishments alive in 2018 that were in age bin 4 (15-19) at any point between 1998 and 2013.
- We also repeat steps 1-5 above restricting the data to establishments alive in 2018 that were in age bin 5 (20-24) at any point between 1998 and 2013.

Columns 2-5 of Tables A.1 and A.2 shows the results. In Table A.1 we include the number of unique establishments in each exercise. Note that data from the same establishment might be exploited in every separate regression if the establishment is observed for at least two periods (in addition to 2018).

Repeating the strategy in steps 1-5 using different starting ages will result in different estimates for the cumulative growth rate through the same age, as shown in Figure A.1. For example, for the average growth rate from ages 5-9 to ages 10-14, we can use the rates from the first panel (with starting ages 0-4) or from the second one (with starting ages 5-9). We average across these results by running the estimated life cycles on dummies for age and dummies for the starting age, separately for formal, informal, and the average plant. Figure 2 in the paper plots the resulting growth rates, which we use in our calibration.

Table A.1: OLS estimates across different starting ages.

	(1) [0-4]	(2) [5-9]	(3) [10-14]	(4) [15-19]	(5) [20-24]
[5-9]	0.253^{***}				
	(0.016)				
[10-14]	0.361^{***}	0.231^{***}			
	(0.024)	(0.005)			
[15-19]	0.440^{***}	0.320^{***}	0.208^{***}		
	(0.038)	(0.007)	(0.006)		
[20-24]	0.464^{***}	0.366^{***}	0.269^{***}	0.170^{***}	
	(0.046)	(0.010)	(0.009)	(0.026)	
[25-29]		0.398^{***}	0.334^{***}	0.222^{***}	0.185^{***}
		(0.015)	(0.014)	(0.042)	(0.010)
[30-34]			0.394^{***}	0.228^{***}	0.256^{***}
			(0.022)	(0.067)	(0.017)
[35-39]				0.311^{***}	0.310^{***}
				(0.099)	(0.026)
Formal	-0.023	-0.020^{***}	-0.014^{***}	-0.022	-0.006
	(0.014)	(0.004)	(0.005)	(0.020)	(0.007)
[5-9]xFormal	0.264^{***}				
	(0.023)				
[10-14]xFormal	0.382^{***}	0.151^{***}			
	(0.028)	(0.006)			
[15-19]xFormal	0.311^{***}	0.172^{***}	0.128^{***}		
	(0.037)	(0.008)	(0.008)		
[20-24]xFormal	0.414^{***}	0.208^{***}	0.167^{***}	0.131^{***}	
	(0.048)	(0.009)	(0.009)	(0.033)	
[25-29]xFormal		0.316^{***}	0.173^{***}	0.236^{***}	0.142^{***}
		(0.016)	(0.013)	(0.044)	(0.012)
[30-34]xFormal			0.188^{***}	0.198^{***}	0.153^{***}
			(0.025)	(0.070)	(0.017)
[35-39]xFormal				0.213^{*}	0.150^{***}
				(0.115)	(0.026)
Sector	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes
Synth. panel	Yes	Yes	Yes	Yes	Yes
Observations	6,004,576	4,458,683	3,045,703	2,047,579	1,227,890
Unique obs.	3,006,254	2,088,920	1,450,693	1,042,205	706,722
R^2	0.000	0.004	0.003	0.000	0.002

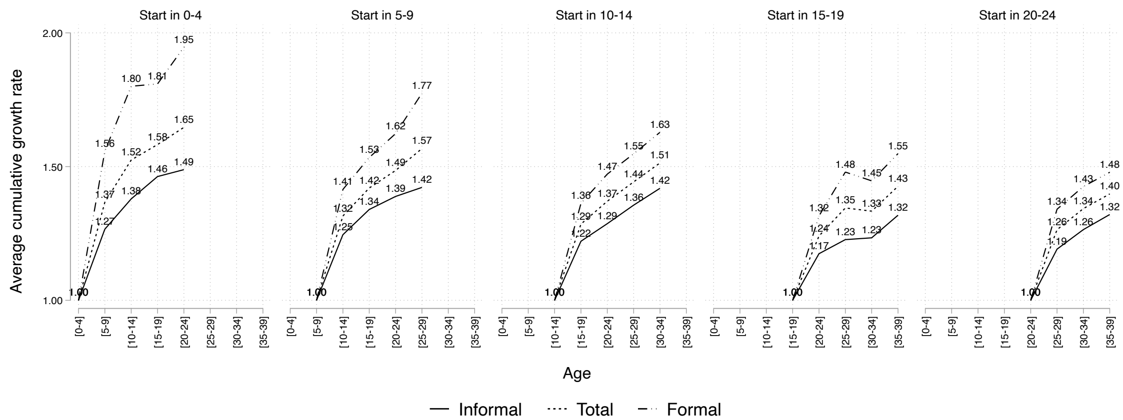
Notes: The dependent variable is the growth in employment from the base category (base=1). The analysis only considers plants in the 2018 cross-section (where we observe their formality status). We include controls for the 2-digit sector (SCIAN), cohort fixed effects, and a dummy for whether the panel identifier is synthetic. This dummy is 1 for observations in the period 1998-2008, and 0 thereafter. Formal establishments are registered with SAT, IMSS, or both. Establishments that are not formal are informal. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Marginal effects of the age fixed effects for formal and informal establishments.

	(1) [0-4]	(2) [5-9]	(3) [10-14]	(4) [15-19]	(5) [20-24]
[0-4]xInformal	0.951				
[0-4]xFormal	0.928				
[5-9]xInformal	1.203	0.942			
[5-9]xFormal	1.445	0.922			
[10-14]xInformal	1.311	1.173	0.940		
[10-14]xFormal	1.670	1.304	0.925		
[15-19]xInformal	1.391	1.262	1.147	0.976	
[15-19]xFormal	1.678	1.414	1.261	0.954	
[20-24]xInformal	1.415	1.308	1.209	1.146	0.967
[20-24]xFormal	1.806	1.496	1.362	1.254	0.961
[25-29]xInformal		1.340	1.274	1.197	1.152
[25-29]xFormal		1.635	1.432	1.411	1.288
[30-35]xInformal			1.333	1.204	1.223
[30-35]xFormal			1.507	1.379	1.370
[35-39]xInformal				1.286	1.277
[35-39]xFormal				1.477	1.421

Notes: Marginal effects of the interaction of the age fixed effects with the binary indicator for whether the establishment is formal. Formal establishments are registered with SAT, IMSS, or both. Establishments that are not formal are informal. The analysis only considers plants in the 2018 cross-section (where we observe their formality status). We include controls for the 2-digit sector (SCIAN), cohort fixed effects, and a dummy for whether the panel identifier is synthetic. This dummy is 1 for observations in the period 1998-2008, and 0 thereafter.

Figure A.1: Estimated growth rates through different ages for different starting bins.



Notes: Marginal effects of the interaction of the age fixed effects with the binary indicator for whether the establishment is formal. The effects are rescaled by dividing by the value at ages [0-4]. Formal establishments are registered with SAT, IMSS, or both. Establishments that are not formal are informal. The analysis only considers plants in the 2018 cross-section (where we observe their formality status). We include controls for the 2-digit sector (SCIAN), cohort fixed effects, and a dummy for whether the panel identifier is synthetic. This dummy is 1 for observations in the period 1998-2008, and 0 thereafter.

B. Additional tables and figures

Table B.3: The establishment and employment size distribution in the census. 2018.

Size	Total number of establishments	Share	Total number of workers	Share of employment	Average size
1	1,818,816	42.26%	1,818,816	8.88%	1.00
2	1,223,455	28.42%	2,446,910	11.95%	2.00
3-5	848,306	19.71%	3,029,170	14.79%	3.57
6-10	226,043	5.25%	1,655,406	8.08%	7.32
11-20	88,567	2.06%	1,274,736	6.22%	14.39
21-50	51,263	1.19%	1,606,644	7.84%	31.34
51-100	16,745	0.39%	1,190,401	5.81%	71.09
101-250	12,411	0.29%	1,908,394	9.32%	153.77
251-500	4,296	0.10%	1,513,206	7.39%	352.24
501-1000	1,966	0.05%	1,359,090	6.64%	691.30
1001+	1,241	0.03%	2,678,432	13.08%	2,158.29
Total	4,293,109	100%	20,481,205	100%	4.77

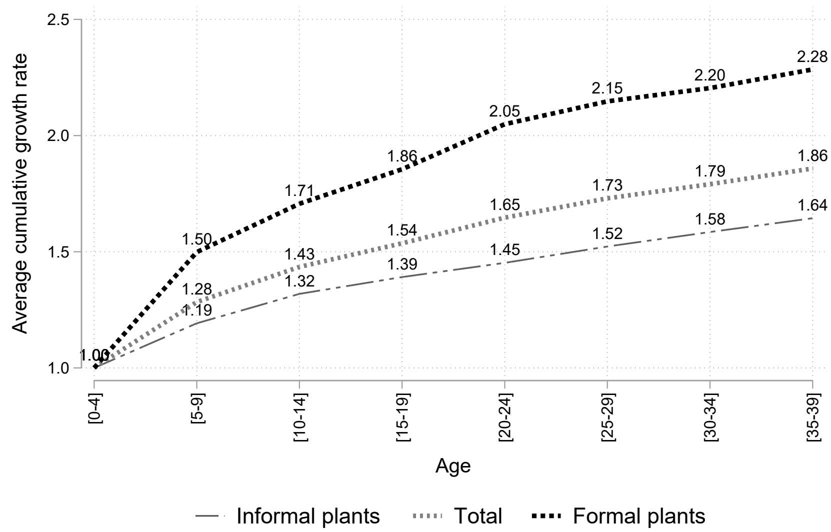
Table B.4: Correlation between formality on the extensive margin and formality on the intensive margin. 2018.

		Extensive margin		
		Formal	**Informal**	**Total**
Intensive margin	**Formal**	8.93	0.86	9.80
	Informal	24.64	65.57	90.20
	Total	33.57	66.43	100.00

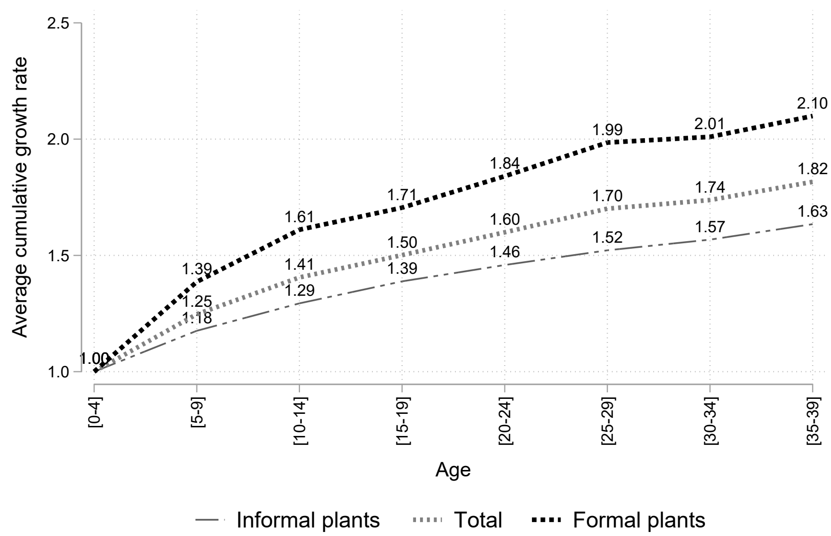
Notes: An establishment is formal on the extensive margin if they are registered with SAT, IMSS, or both. Following Busso et al. (2012) and Levy (2018), an establishment is formal on the intensive margin if they employ at least one salaried worker and pay any positive amount in mandatory social security contributions.

Figure B.2: Cumulative growth rates in manufacturing and services.

(a) **Manufacturing**



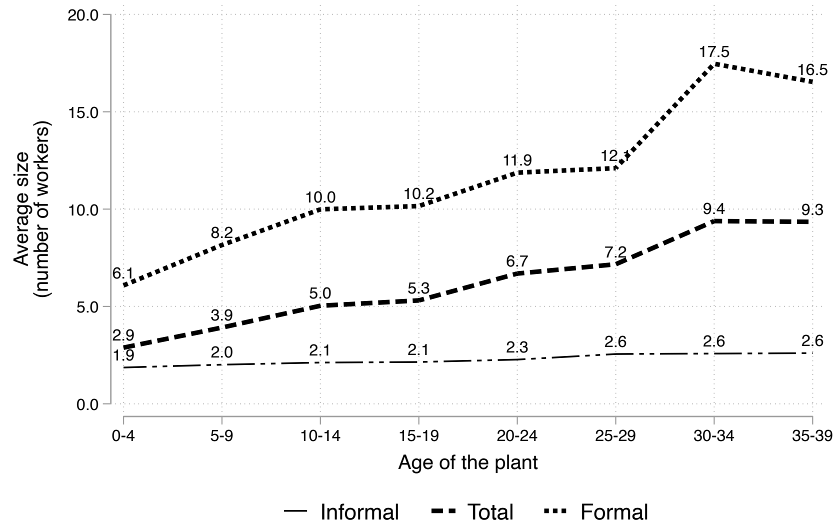
(b) **Services**



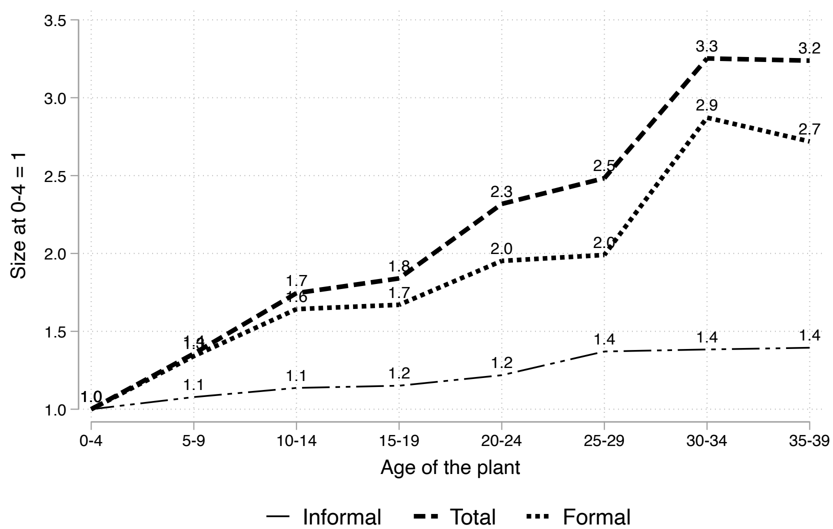
Notes: Formal establishments are registered with SAT, IMSS, or both. Establishments that are not formal are informal. To estimate the growth rates during the life cycle, we follow the methodology described in Section A of the appendix but restricting the sample to manufacturing in panel a and services in panel b.

Figure B.3: Cross-section estimates of the life cycle of plants. 2018.

(a) Average number of workers

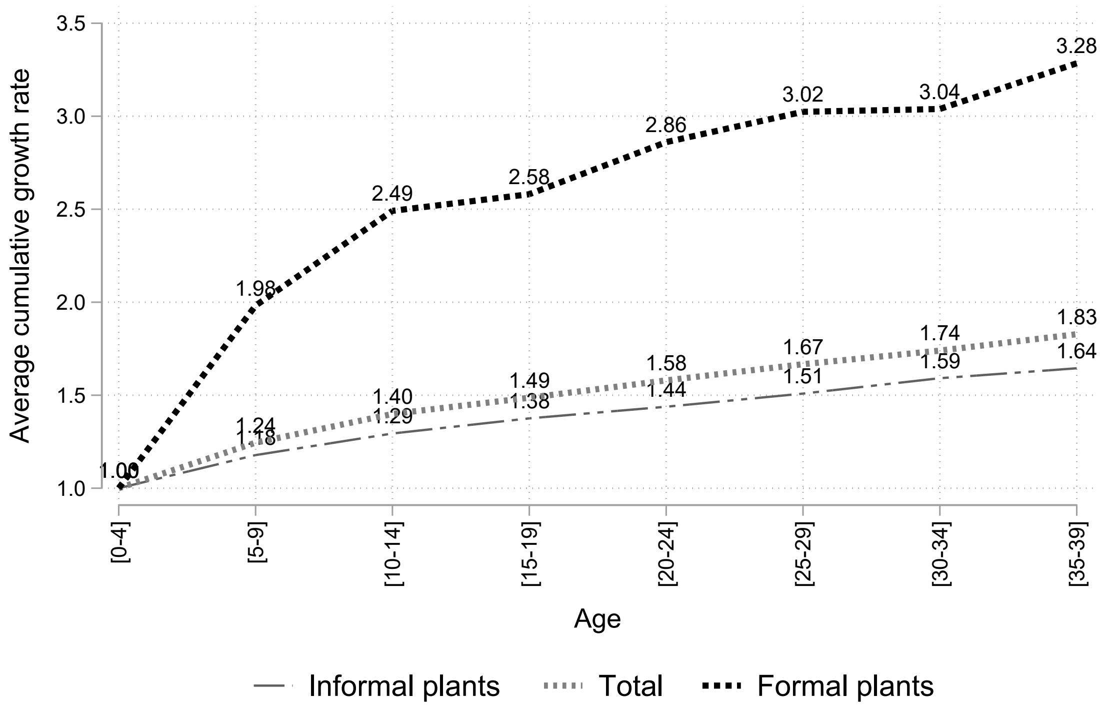


(b) Growth rates



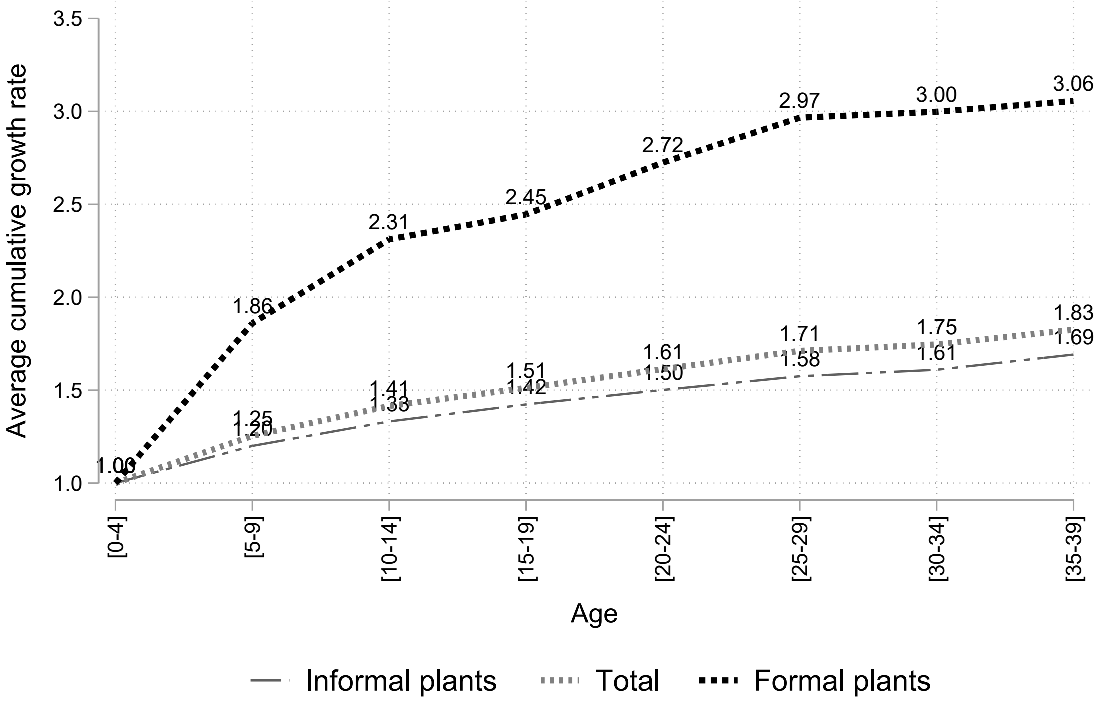
Notes: The top panel shows simple averages of the size of formal and informal plants in each age bin in the 2018 cross-section. The bottom panel computes growth rates using the data from panel a.

Figure B.4: Estimates for the life cycle of formal and informal plants using a strict definition of informality.



Notes: Formal establishments are registered with both SAT and IMSS. Informal establishments are not registered neither with SAT nor with IMSS (they answered no to both questions). To estimate the growth rates during the life cycle, we follow the methodology described in Section A of the appendix.

Figure B.5: Estimates for the life cycle of formal and informal plants using the definition of informality on the intensive margin from Busso et al. (2012) and Levy (2018).



Notes: In this figure we apply the definition of informality from Busso et al. (2012) and Levy (2018). Informal establishments either: (a) employ non-salaried workers only; (b) employ salaried workers only or a mixture of salaried and non-salaried, but in any case, they fully evade their fiscal obligations in terms of their social security payments. Formal establishments employ at least one salaried worker and can also employ non-salaried workers; additionally they pay any positive amount of their social security contributions. To estimate the growth rates during the life cycle, we follow the methodology described in Section A of the appendix.

C. First Order Conditions for Factor Inputs and Equilibrium Definition

C.1 Factor Input Decisions of Formal Entrepreneurs

$$k^F(z) = [A(1-\tau)\gamma]^{\frac{1}{1-\gamma}} \left(\frac{\alpha}{R}\right)^{\frac{(\alpha-1)\gamma+1}{1-\gamma}} \left(\frac{w}{1-\alpha}\right)^{\frac{(\alpha-1)\gamma}{1-\gamma}} z \tag{C.1}$$

$$n^F(z) = [A(1-\tau)\gamma]^{\frac{1}{1-\gamma}} \left(\frac{\alpha}{R}\right)^{\frac{\alpha\gamma}{1-\gamma}} \left(\frac{w}{1-\alpha}\right)^{\frac{\alpha\gamma-1}{1-\gamma}} z \tag{C.2}$$

C.2 Factor Input Decisions of Informal Entrepreneur

Unconstraint:

$$k^I(z) = [A\gamma]^{\frac{1}{1-\gamma}} \left(\frac{\alpha}{R}\right)^{\frac{(\alpha-1)\gamma+1}{1-\gamma}} \left(\frac{w}{1-\alpha}\right)^{\frac{(\alpha-1)\gamma}{1-\gamma}} z \tag{C.3}$$

$$n^I(z) = [A\gamma]^{\frac{1}{1-\gamma}} \left(\frac{\alpha}{R}\right)^{\frac{\alpha\gamma}{1-\gamma}} \left(\frac{w}{1-\alpha}\right)^{\frac{\alpha\gamma-1}{1-\gamma}} z \tag{C.4}$$

Constraint

$$n^I(z) = \left[\frac{A\gamma(1-\alpha)b^{\alpha\gamma}}{w}\right]^{\frac{1}{1-\gamma(1-\alpha)}} z^{\frac{1-\gamma}{1-\gamma(1-\alpha)}} \tag{C.5}$$

C.3 Stationary Equilibrium

Given factor prices, w and R, distortions and enforcement levels, τ, b and $\bar{\phi}$, in a stationary equilibrium

1. Entrepreneurs and workers solve their problem which is consistent with their occupational choices

2. All agents solve their occupation choice problem such that choices are consistent with the solution of enterpreneurs and workers' problems.

3. All markets clear.

   - Labor market:

$$\sum_{j=1}^{J_R-1} \int_{z\in S_{W,ne}} (1-\bar{\phi})dG(z) + \sum_{j=1}^{J_R-1} \int_{z\in S_{W,e}} \bar{\phi}dG(z) = \sum_{j=1}^{J_R-1} \int_{z\in S_{I,ne}} (1-\bar{\phi})n^I(z)dG(z)$$
$$+ \sum_{j=1}^{J_R-1} \int_{z\in SF,ne} (1-\bar{\phi})n^F(z)dG(z) + \sum_{j=1}^{J_R-1} \int_{z\in S_{F,e}} \bar{\phi}n^F(z)dG(z)$$

where $S_{W,ne}$ and $S_{W,e}$ are the set of workers among agents who are assigned to a tax official and who are not assigned to a tax official respectively. Moreover, $S_{I,ne}$ is the set of informal enterpreneurs and $S_{F,ne}$ and $S_{F,e}$ stand for the set of formal enterpreneuers who are not assigned to a tax official and who are assigned to a tax official respectively. Hence, the first and second terms left hand side of the labor market clearing condition is the total supply of labors from agents who are not assigned to tax official and agents who are assigned to tax official respectively. The right hand side of the labor market clearing condition is the total demand for labor by informal and formal entrepreneurs: in the right-hand side, the first term is the demand for labor from informal plants (those that don't meet the tax official); the second term is the demand for labor from agents who don't meet the tax official but optimally choose to become formal; and the third item is the demand for labor from entrepreneurs who meet the tax official and have to operate in the formal sector.

- Capital market

$$\sum_{j=1}^{J_R-1}\int_{z\in S_{I,ne}}(1-\bar{\phi})k^I(z)dG(z)+\sum_{j=1}^{J_R-1}\int_{z\in S_{F,ne}}(1-\bar{\phi})k^F(z)dG(z)+$$
$$\sum_{j=1}^{J_R-1}\int_{z\in S_{F,e}}\bar{\phi}k^F(z)dG(z)=\sum_{j=1}^{J-1}\int_{z\in S_{W,ne}}(1-\bar{\phi})a^W(z)dG(z)+\sum_{j=1}^{J-1}\int_{z\in S_{W,e}}\bar{\phi}a^W(z)dG(z)+$$
$$\sum_{j=1}^{J-1}\int_{z\in S_{I,ne}}(1-\bar{\phi})a^I(z)dG(z)+\sum_{j=1}^{J-1}\int_{z\in S_{F,ne}}(1-\bar{\phi})a^F(z)dG(z)+\sum_{j=1}^{J-1}\int_{z\in S_{F,e}}\bar{\phi}a^F(z)dG(z)$$

where $a^W(z)$, $a^I(z)$,$a^F(z)$ denote the saving of a worker, informal and formal entrepreneur with ability z respectively. The left hand-side of the capital market clearing condition is capital demand by entrepreneurs and the right hand-side is the total savings of agents in the economy.

- Goods market clears

$$\sum_{j=1}^{J}\int_{z\in S_{W,ne}}(1-\bar{\phi})[c_j^W(z)+a_j^W(z)]dG(z)+\sum_{j=1}^{J}\int_{z\in S_{I,ne}}(1-\bar{\phi})[c^I(z)+a^I(z)+x^I(z)]dG(z)+$$
$$\sum_{j=1}^{J}\int_{z\in S_{F,ne}}(1-\bar{\phi})[c_j^F(z)+a_j^F(z)+x_j^F(z)]dG(z)+\sum_{j=1}^{J}\int_{z\in S_{W,e}}\bar{\phi}[c_j^W(z)+a_j^W(z)]dG(z)$$
$$+\sum_{j=1}^{J}\int_{z\in S_{F,e}}\bar{\phi}[c_j^F(z)+a_j^F(z)+x_j^F(z)]dG(z)=\sum_{j=1}^{J_R-1}\int_{z\in S_{I,ne}}(1-\bar{\phi})y^I(z)dG(z)+$$
$$\sum_{j=1}^{J_R-1}\int_{z\in S_{F,ne}}(1-\bar{\phi})y^F(z)dG(z)+\sum_{j=1}^{J_R-1}\int_{z\in S_{F,e}}\bar{\phi}y^F(z)dG(z)+(1-\delta)K$$

where K denotes the capital stock.