(In)formal Growth: Wage Dynamics in Developing Economies*

Santiago Franco
University of Chicago

Jose M. Quintero
University of Chicago

January, 2024

Preliminary. Click here for the latest version.

Abstract

Labor informality is pervasive in developing economies. In this paper, we investigate the interconnection between informal labor, human capital accumulation, and economic growth. How do informal labor markets affect human capital accumulation, and vice versa? What are the aggregate effects of this interaction on growth and welfare? Using panel data from Chile and Colombia, we explore the dynamics of the formal and informal sectors by documenting two new empirical facts. First, wages for formal workers increase significantly more over the life cycle than wages for informal workers. Second, a substantial portion of this formal wage premium is attributable to workers' skill-based sorting. To rationalize these patterns, we build an endogenous growth model where heterogeneous workers sort into formal and informal labor markets based on their potential earnings. Worker's human capital increases over their life cycle through interactions with other workers. In equilibrium, more knowledgeable workers sort into the formal sector, and the growth rate of the economy is determined by the rate at which all workers meet more knowledgeable formal workers. We structurally estimate the parameters of the model and use it to quantify the effect of formalization policies. We find that policies that decrease the cost of operating formally are more effective in reducing the size of the informal sector compared to policies that increase the cost of producing informally. However, both types of policies have adverse effects on economic growth by lowering the quality of interactions of more skilled workers.

*Franco: sfranco@uchicago.edu. Quintero: jmquintero925@uchicago.edu. We are grateful to Santiago Caicedo, Levi Crews, Chang-Tai Hsieh, Erik Hurst, Ben Moll, Aleksei Oskolkov, Esteban Rossi-Hansberg, Pascual Restrepo, and Roman David Zarate for insightful discussions. We thank DANE for facilitating the access to the Colombian data and the Subsecretaria de Prevision Social of Chile for the access to the Chilean data. All errors are our own.

1 Introduction

Informal labor markets are prevalent in developing economies. In Latin America, between 35% and 80% of workers do not have a formal labor contract, and the informal sector accounts for 40% of the national GDP.[1] Despite the coexistence of formal and informal firms within the same industries, producing similar products, wages for informal and formal workers vary widely.[2] Understanding the determinants of these wage differentials is crucial because they can be important factors shaping overall income inequality, and governments spend vast resources every year to decrease informality. Importantly, when wage differences reflect differences in human capital between workers, wage dynamics over the life cycle of formal and informal workers can have key implications for economic growth and long-term welfare. How do informal labor markets affect the formation of human capital? How does the stock of talent in an economy determine the size of the informal sector? How does this loop feed into economic growth and welfare?

This paper addresses these questions with three main contributions. First, utilizing worker panel data from Chile and Colombia, we document two new empirical findings on wage dynamics for formal and informal workers. By exploiting the panel structure of our data to account for unobserved worker heterogeneity, we find that i) a significant portion of the formal wage premium is attributable to workers sorting based on skill, and ii) formal workers experience higher wage growth throughout their life cycle. Second, to rationalize these findings, we develop a heterogeneous agent endogenous growth model with formal and informal labor markets. The model explains wage differentials between formal and informal sectors through workers' sorting based on skill and differential human capital accumulation over the life cycle. Third, we estimate the model and use it to quantify the aggregate effects of two different types of formalization policies on the size of the informal labor market, the growth rate of the economy, and welfare.

In the first part of the paper, we examine wage differentials between formal and informal workers using rolling surveys in Chile (EPS) and Colombia (ELCO).[3] Each survey allows us to classify workers as formal or informal by examining their affiliation with social security and pension funds. In addition, the surveys track individual workers across different years. Therefore, compared to previous studies, the structure of our data allows us to study wage changes for the same worker across the formal and informal sectors and analyze wage dynamics for particular workers over the life cycle. Exploiting these data characteristics, we document two new empirical facts.

We first document that wages for formal workers grow significantly more over their life cycle compared to wages for informal workers. In Chile, the wages of formal workers increase by 80% over the life cycle, while the wages of informal workers only increase by 40%. Similarly, wages for formal workers in Colombia increase by 50% throughout the life cycle, while wages for informal workers only increase by 10%. We leverage our panel structure to show that this pattern is robust even when accounting for observable characteristics and time-invariant unobserved heterogeneity. Furthermore, we used the richness of our Chilean data to demonstrate that this pattern persists even

[1]See Perry et al. (2007).
[2]See Ulyssea (2020b) for a comprehensive review of informal firms and workers.
[3]EPS: *Encuesta de Proteccion Social.* ELCO: *Encuesta Longitudinal de Colombia.*

when considering the number of jobs and experience, and it is not solely explained by differences in the job ladder. Consequently, we assert that the divergence in wage growth for both sectors is attributed to differences in the formation of human capital within each sector.[4]

Secondly, consistent with previous evidence in other developing countries, we observe a significant wage premium for formal workers compared to their informal counterparts. Controlling for observable characteristics, formal workers in Chile earn wages 34% higher than their informal counterparts, while in Colombia, this wage premium increases to 44%. These findings align with the results of Bargain & Kwenda (2014) and Ulyssea (2018). However, we discover that a substantial portion of this wage gap can be attributed to high-skilled workers sorting into the formal sector. Once we account for workers' unobserved time-invariant characteristics, the formal wage premium reduces to 21% in Chile and 15% in Colombia. Additionally, Chilean data provide a complete labor history for each worker, allowing us to control for experience in both formal and informal sectors separately for each individual. Consequently, our estimates account for the returns to experience in both sectors, even when the skills required for different sectors are not perfectly transferable.

In the second part of the paper, to rationalize our empirical findings, we propose an endogenous growth model with heterogeneous workers and informal labor markets. Workers differ in their level of human capital (skill) and choose between operating in the formal and informal sectors. To operate formally, workers incur a fixed cost and are also subject to labor taxes. While we do not interpret this cost purely as monetary, it aims to encompass variable costs of the formal sector, such as social security and pension contributions, as well as fixed costs, like registering with government authorities. In contrast, informal workers avoid labor taxes and fixed costs but are subject to potential fines imposed by the government, which increase with the level of production. As production levels increase, the likelihood of government detection also rises.[5] The government rebates labor taxes as lump-sum transfers to all workers in the economy, regardless of their formalization status.[6]

Wages increase over the life cycle as workers accumulate human capital by learning from more skilled workers. Workers interact with other workers at a common exogenous rate. However, building on the observations of Jarosch et al. (2021), conditional on a meeting, we allow the probability of an informal worker interacting with a formal worker to differ from that of a formal worker interacting with another formal worker. Therefore, workers choose to operate in each of the sectors not only based on the static returns that depend on their current skill and labor market regulations but also on the potential learning opportunities that determine future wage paths. The model is tractable enough to imply a unique cutoff in the productivity distribution, for which workers with higher skills sort into the formal sector and workers with lower skills into the informal sector. In equilibrium, the economy's growth rate in a Balanced Growth Path is determined by the dispersion of productivity in the economy and a weighted average of meeting rates with the most knowledgeable agents.

[4] See Lagakos et al. (2018) for a discussion of potential mechanisms that explain different wage growth patterns over the life cycle.

[5] This can also be interpreted as expenses that informal workers must bear to avoid detection by authorities.

[6] This structure is consistent with social security programs in developing economies, in which high-income households subsidize access for low-income households.

The size of the informal labor market has implications for growth and welfare. On the one hand, a large informal sector results in high output losses for the economy, as more resources are used to conceal operations from the government. Additionally, if learning frictions are high enough, a larger labor informality diminishes human capital improvements for the least-skilled workers who sort into the informal sector. On the other hand, a smaller informal sector has two effects on the quality of interactions in the formal sector. First, it increases the expected human capital improvements for workers who formalize, as they have better opportunities to learn from more skilled workers. Second, it decreases the quality of human capital in the formal sector, as the marginal agent that formalizes is less productive. With less skilled workers crowding out meetings in the formal sector, it becomes harder for all formal workers to meet more skilled counterparts. The relative strength of these two forces determines whether reductions in labor informality have a positive or negative effect on growth and welfare. Therefore, depending on parameter values, reductions in the informal sector might actually decrease growth and welfare in the long run.

In the final section of the paper, we estimate the model's parameters to perform counterfactual exercises. We employ a Simulated Method of Moments (SMM) approach to estimate most of the model's parameters, focusing on Colombia. Leveraging the model structures, we infer the parameters that determine workers' dynamic returns from estimated transition matrices between the formal and informal sectors and the aggregate growth rate. These parameters include the worker's meeting rates, the meeting probabilities for formal and informal workers, and the tail of the overall initial productivity distribution. Moreover, we infer parameters that determine workers' static returns from the estimated wage paths over the life cycle for formal and informal workers and the share of informal workers. These parameters encompass the formal sector fixed cost, labor tax, and the informal sector production cost parameters. The remaining parameters are estimated directly from the data using a Generalized Method of Moments (GMM).

We use our estimated model to assess the effects of two types of formalization policies. On the one hand, we consider a policy that augments the prosecution of informal activity, increasing the cost of operating informally (sticks). On the other hand, we consider a policy that decreases formalization costs (carrots). We find that the latter policy is significantly more effective at reducing labor informality. Decreasing registering cost in the formal sector by 50% decreases informality in the long run from 63% to 48%. In contrast, increasing auditing efforts by the same amount has almost negligible effects, decreasing the share of informal labor only to 60%.

Surprisingly, we find that both types of formalization policies can have an adverse effect on growth and welfare. Since both policies move less skilled workers into the formal sector, the stock of human capital in the formal sector decreases, lowering the quality of interactions in this sector. This crowding-out effect dominates the potential gains from human capital improvements for the newly formalized workers, leading to a decrease in the growth rate and welfare losses. Nevertheless, we find that decreasing the formal sector registering costs has a milder negative impact on growth and welfare than increasing the informality costs. These results suggest that the economy would benefit in the long run from more segmented markets, where high-skill workers initiate learning among themselves in the formal sector, expanding the knowledge frontier. Subsequently, the rest of the less skilled informal workers catch up.

Related Literature. Our work contributes to three strands of the literature. The first strand involves the study of knowledge diffusion and economic growth. Our model aligns with the concept of a mean-field game, as introduced by Lasry & Lions (2007), with the productivity distribution serving as a state of the economy. Previous studies, such as those by Lucas Jr (2009), Lucas & Moll (2014), and Perla & Tonetti (2014), have developed frameworks in which knowledge diffusion occurs through interactions or imitation. We extend this work by introducing multiple sectors. In our model, workers endogenously sort into formal and informal sectors based on their static returns and learning opportunities, following the approach of Akcigit et al. (2018). Moreover, we allow learning opportunities to vary between sectors, capturing the idea that more productive colleagues imply greater spillovers, as demonstrated by Jarosch et al. (2021). Consequently, our model establishes a productivity cutoff that effectively sorts workers into formal and informal sectors, akin to the framework presented by Perla et al. (2021).

Secondly, our work contributes to the literature that examines wage dynamics in developing economies. The study of wage profiles throughout the life cycle in developing economies is limited, and our research takes a step toward understanding frictional labor markets in these settings. Lagakos et al. (2018) is among the first to explore wages across the life cycle for both developing and developed economies. They find steeper wage profiles throughout the life cycle for developed economies and discuss potential mechanisms driving this pattern. In contrast, our work identifies one friction that contributes to this phenomenon. We demonstrate that wages grow faster in the formal sector, and the presence of a large informal sector suppresses average wage growth. Beneath the differing wage profiles for formal and informal workers, we develop a theory explaining variations in human capital accumulation. Our findings align with Engbom (2022), where higher contact rates lead to increased human capital and, consequently, higher wage growth. However, in our framework, the accumulation of human capital is the primary driver of wage growth throughout the life cycle, consistent with the insights from Ben-Porath (1967) and Heckman et al. (1998).

Finally, our work contributes to the literature on informal labor markets and their impact on economic performance. There is a growing body of literature examining labor informality from a macro perspective, with notable contributions from Ulyssea (2018), Dix-Carneiro et al. (2021), and Meghir et al. (2015). More closely related to our work, Bobba et al. (2022) investigates the relationship between informal markets and human capital investments. However, these studies primarily focus on the static effects of informality. In contrast, we forefront the dynamics and explore how informal markets not only affect aggregate production but also influence economic growth. Similarly, Lopez Garcia (2015) examines how informal markets shape incentives to invest in human capital, yet their work overlooks the feedback effect of human capital in determining the size of the informal sector. To address this gap, we build a general equilibrium model where the stock of human capital and the share of informality are endogenously determined. Lastly, a recent group of articles investigates the effects of informality on economic growth, including works by Akcigit et al. (2022) and Lopez-Martin (2019). However, these papers approach informality from the firm perspective and do not delve into the intricacies of human capital and economic performance, as emphasized by Manuelli & Seshadri (2014).[7]

[7] Ulyssea (2020a) highlights the need to study workers dynamic decision regarding formality status.

The rest of the paper follows this structure. Section 2 describes the institutional settings, presents our empirical findings, and discusses the data. Section 3 introduces our model and outlines the equilibrium. Section 4 details our estimation strategy and provides an overview of our sources of identification. In Section 5, we present the results of our counterfactual exercises. Finally, Section 6 offers concluding remarks.

2 Empirical Analysis

Defining labor informality has been one of the main challenges for the literature studying these labor markets. In this section, we start by providing an overview of the labor regulations in both Colombia and Chile, which will serve as motivation for our definition. Following that, we offer a brief overview of the data for both countries and specify the variables of interest for our empirical exercises. Finally, we present our key empirical findings.

2.1 Institutional Details

Measuring informality is a challenge because there are multiple ways to define an informal worker. On the one hand, one could define an informal worker as anyone employed by a firm not registered with tax authorities. However, this approach would not be comprehensive of all the informal activity in the economy, since formally registered firms will also have a mix of formal and informal workers. For example, Samaniego De La Parra & Fernández Bujanda (2020) estimates that 26% of the formal workers of the economy are employed by a formally registered establishment. Instead, we define an informal worker as any employee whose contract does not comply with local labor regulations. In the Chilean case, labor regulations require workers to be affiliated with a pension fund and contribute 10% of their taxable income. In particular, the following articles from *Ley 3500* (1980) broadly summarize the requirements to be categorized as a formal worker.

1. **Affiliation.** At the start of any labor relation, any non-affiliated worker generates automatic affiliation to the (pension) System and the obligation to contribute to a pension fund... Membership is the legal relationship between a worker and the *Sistema de Pensiones de Vejez, Invalidez y Sobrevivencia* that gives rise to the rights and obligations established by law... Each worker, even if he or she provides services to more than one employer, may only contribute to one pension fund (Article 2, Law 3500 of 1980).
2. **Contribution.** Workers affiliated with the (pension) system, under 65 years of age if they are men and under 60 years of age if they are women, will be required to contribute 10% of their salaries and income to their individual pension account (Article 17, Law 3500 of 1980).

For the Colombian case, labor regulation is even more stringent, as it requires that workers are affiliated not only with pension funds but also to have medical insurance. In particular, the regulations for Colombia are presented in *Ley 100* (1993):

1. **Pension Fund Affiliation.** Individuals affiliated with the General Pension System: *(i)* Mandatory Affiliation: All individuals connected through an employment contract or as

public servants, except for the exceptions stipulated in this law. Likewise, population groups that, due to their characteristics or socio-economic conditions, are eligible to be beneficiaries of subsidies through the *Fondo de Solidaridad Pensional*, in accordance with budgetary availability (Article 15, Law 100 of 1993).

2. **Pension Fund Contribution.** During the validity of the employment contract, affiliates and employers must make mandatory contributions to the General Pension System Contributions must be made by affiliates and employers based on the salary they earn. Except as provided in Article 64 of this Law, the obligation to contribute ceases when the affiliate meets the requirements to access the minimum old-age pension or when the affiliate retires due to disability or early retirement (Article 17, Law 100 of 1993).

3. **Medical Insurance Affiliation.** Affiliation to the General Social Security System in Health is mandatory for all residents of Colombia. Consequently, it is the responsibility of every employer to affiliate their workers to this system, and the State is tasked with facilitating affiliation for those without a connection to any employer or the financial capacity to pay (Article 153, Law 100 of 1993.).

4. **Medical Insurance Contribution.** Duties of Employers. As members of the General Social Security System in Health, employers, regardless of the entity or institution on behalf of which they engage workers, must: *(i)* Enroll all individuals with any form of labor connection, verbal or written, temporary or permanent, in a Health Promoting Entity. Collective affiliation should, under no circumstances, limit the worker's freedom of choice regarding the Health Promoting Entity they prefer to affiliate with, in accordance with the regulations. *(ii)* In accordance with Article 22 of this Law, contribute to the financing of the General Social Security System in Health through actions such as the following: *(ii.a)* Promptly pay the contributions owed, in accordance with Article 204. *(ii.b)* Deduct the contributions owed by the workers in their service from their labor income... (Article 161, Law 100 of 1993).

We abstract from self-employed workers in this study, as the definition of informality for this group of individuals requires additional considerations as established by the labor codes of each country. In addition, we claim that the dynamics for this particular type of worker are very different than for employees working at a firm.

2.2 Data

Our data sources consist of two rolling panels with information on worker demographics, contributions to the Social Security system, and wages. For Colombia, we utilize the *Encuesta Longitudinal Colombiana* (ELCO). The ELCO is a longitudinal survey that tracks approximately 10,000 Colombian households in rural and urban areas every three years. The objective of the survey is to observe and analyze the same households over a 12-year span, with the survey currently covering three waves: 2010, 2013, and 2016. Importantly, ELCO follows a set of individuals over time, allowing the creation of a panel data set that provides a dynamic perspective on the social and economic changes experienced by these individuals and their households. We divide our sample into two for our analysis: *(i)* Cross-sectional sample, where we combine repeated cross-sectional observations with the panel observations, and *(ii)* Panel sample, a subset of our cross-sectional sample, considering only individuals tracked over the three waves. For our analysis, we focus on urban households and exclude self-employed workers from the sample. Table A-1 in the Appendix presents descriptive statistics for both samples.

Our data source for Chile is the *Encuesta de Protección Social* (EPS). This is a comprehensive longitudinal survey designed to gather essential data on various aspects of social protection. The survey spans five waves conducted in 2002, 2004, 2006, 2009, and 2015. However, the first wave only considers workers' affiliation to a pension fund, and thus we exclude it from our analysis because it does not account for informal workers. The survey maintains a nationally representative sample of 14,045 individuals across the four waves, with remarkably low attrition rates. It encompasses a nationally representative sample that provides insight into the diverse experiences and conditions of individuals in different regions and demographic groups. Participants are asked to provide detailed information on various aspects of their lives, including family history, educational history, and labor activities. This wealth of information allows for nuanced analyses and a comprehensive understanding of participants' life trajectories.

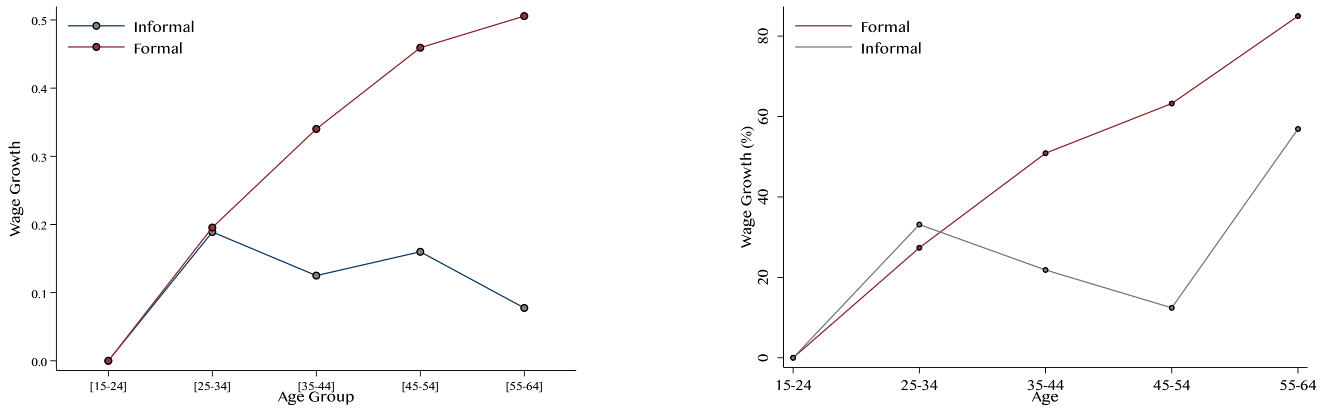
During the initial wave, participants are asked to report their family history, full educational history, and elaborate on their labor activities dating back to 1980. This includes information on the types of job held, the number of hours worked, whether they contributed to social security during those employments, and their labor status (e.g., whether they were employed by a firm or self-employed). The survey captures data related to participants' contributions to social security during their various employments, providing insights into the extent to which individuals have been attached to the informal sector over the years. In addition, participants are asked to detail the nature of their employment, specifying the type of job performed and their employment status (whether they worked in a formal firm or were self-employed). As before, we divided our sample into two cross-sectional and panel samples, and drop self-employed workers. Table A-2 presents descriptive statistics for the relevant variables.

Both data sets provide information about the social security contribution, and hence it allows us to classify workers as either formal or informal. For the Colombian data, we can only observe the current formality status, but we do not observe any transition between sectors within each wave. Regardless, we construct a panel of workers and exploit the variation of movers between the formal and informal sectors to estimate the returns of informality across the life cycle. The Chilean data does not have this limitation, as we have access to all the labor history. This allows us to go beyond and calculate the returns to informal and formal experience in both sectors.

2.3 Stylized Facts

In this section, we present two novel stylized facts that will motivate the mechanisms in our model. First, we demonstrate that formal workers experience higher wage growth over the life cycle compared to informal workers. Second, we show that a significant portion of the formal wage premium is explained by sorting. These facts, to our knowledge, have not been documented before and offer new insights into wage dynamics in developing economies.

Figure 1: Wages over the life-cycle: raw data.



(a) Colombia

(b) Chile

Notes: Figure 1(a) displays wage paths for formal and informal workers over the life cycle in Colombia. Figure 1(b) displays wage paths for formal and informal workers over the life cycle in Chile. Formal worker's wage paths in both figures correspond to the estimates of $\beta_a + \gamma_a$ in equation (1). Informal worker's wage paths in both figures correspond to the estimates of β_a in equation (1). Estimates are obtained without including controls X_{it} and worker's fixed-effects, δ_i. Growth rates are relative to entry-level wages: wages for workers between 15 - 24.

2.3.1 Wages over the life-cycle

The wage gap between formal and informal workers has been extensively studied in the literature[8]. We differ from previous studies by studying the differences in wage growth for formal and informal workers over the life cycle. To do so, consider the following specification.

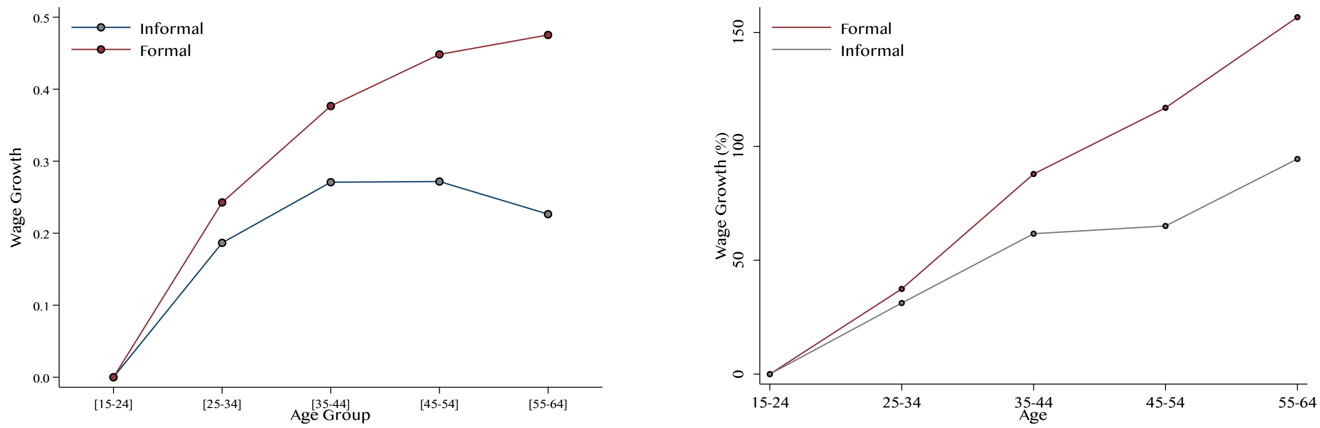
$$\log w_{it} = \beta_0 + \beta_1 \text{Formal}_{it} + \sum_{a \in \mathcal{A}} \left\{ \beta_a \mathbb{1}_{\{a_{it} \in a\}} + \gamma_a \text{Formal}_{it} \times \mathbb{1}_{\{a_{it} \in a\}} \right\} + \Gamma X_{it} + \delta_i + \varepsilon_{it}, \tag{1}$$

where Formal_{it} is a dummy variable equal to one if worker i is formal at time t, a_{it} is the age of a worker i at time t, $\mathcal{A}$ is a partition of ages, X_{it} is a vector of controls, δ_i is an individual fixed effect, and ε_{it} is an idiosyncratic shock normally distributed. Our partition $\mathcal{A}$ consists of intervals of 10 years starting at age 15 and going up to age 65. We discard ages above 65 because it is the retirement age for both Colombia and Chile, and hence the group of workers that remains after 65 suffers from selection. We omit the first age bin from the regression (i.e., from age 15 to 24) so that every coefficient is interpreted relative to the first age bin. Therefore, the coefficient β_a captures the wage premium for informal workers in age bin a relative to ages 15 to 24. Since we are interested in understanding the growth, we consider the trajectory of wages across the life cycle for formal and informal workers relative to the first age bin. In other words, we abstract from the effect of β_1 for formal workers, and for every age bin $a \in \mathcal{A}$, we plot $\beta_a + \gamma_a$. For informal workers, we only plot β_a as the constant term β_0 absorbs the level effect.

Figure 1 displays raw wage paths for both countries. Figures 1(a) and 1(b) present the results

[8] See Bargain & Kwenda (2014)

Figure 2: Wages over the life-cycle: including controls and worker fixed effects.



(a) Colombia **(b)** Chile

Notes: Figure 1(a) displays wage paths for formal and informal workers over the life cycle in Colombia. Figure 1(b) displays wage paths for formal and informal workers over the life cycle in Chile. Formal worker's wage paths in both figures correspond to the estimates of $\beta_a + \gamma_a$ in equation (1). Informal worker's wage paths in both figures correspond to the estimates of β_a in equation (1). Estimates are obtained by including controls X_{it} and worker's fixed-effects, δ_i. Growth rates are relative to entry-level wages: wages for workers between 15 - 24.

of estimates from equation (1) without controlling for workers' observable characteristics, X_{it}, or workers' fixed effects, δ_i. These figures convey two important messages. First, at the beginning of the life cycle, informal workers in both countries experience a steeper wage increase. Second, throughout the entire life cycle, formal workers experience a significantly higher wage increase than workers in the informal sector. In particular, over their careers, formal workers in Colombia experience a wage increase of 50%, while informal workers only see an increase of 10%. Workers in Chile experience a steeper wage increase, with formal workers earning wages 80% higher by the end of their careers, and informal workers experiencing a wage increase of 40%. These results might reflect the fact that formal workers invest more in human capital at the beginning of their careers. Although this has a short-term negative impact on wages, the positive long-run effects are considerable.

Figure 2 displays raw wage paths for both countries, controlling for workers' observable characteristics and workers' fixed effects. For both countries, we still find that wages in the formal sector grow faster than in the informal sector. In Colombia, we observe a net growth of 50%, while in Chile, there is a larger growth of 150%. Our vector of controls, X_{it}, accounts for education level, and hence, the wage growth estimated in Figure 2 does not consider any educational premium the worker may acquire over the life cycle. As a result, we interpret these different trajectories as evidence of differences in human capital accumulation in both the informal and formal sectors.

Admittedly, formal workers can accumulate more human capital on the job for several reasons. However, when this observation is combined with our first fact described in Section 2.3.2, the undoubted peer effects are a relevant mechanism to consider. Specifically, if more skilled workers sort into the formal sector, then interactions and learning from colleagues will benefit more formal

Table 1: Formal premium and Sorting

	log (Wage)				
Panel A: EPS					
Formal Premium	0.470	0.290	0.201	0.196	0.204
	(0.038)	(0.033)	(0.032)	(0.042)	(0.047)
Observations	5932	5932	5932	5932	5932
Panel B: ELCO					
Formal Premium	0.688	0.367	0.143		
	(0.026)	(0.023)	(0.037)		
Observations	5347	5347	5347	-	-
Controls	No	Yes	Yes	Yes	Yes
Worker FE	No	No	Yes	Yes	Yes
Overall experience	No	No	No	Yes	No
Separate experience	No	No	No	No	Yes

Notes: Column 1 depicts the baseline regression coefficients for the EPS in panel A and ELCA in panel B . Column 2 depicts the formal premium controlling for age and education of i at time t. In column 3 we add workers fixed effects. Since the EPS database allows us to follow the full labor history for each worker, we use the experience in each sector as additional controls. Experience is calculated as weekly hours worked in formal or formal jobs. Column 4 in panel A depicts the formal premium after adding experience in the labor market as a control. Column 5 panel A estimates the formal premium when controlling an informal and formal experience control. For all specifications, standard errors are clustered at the worker level.

workers. In other words, the learning opportunities for informal workers are lower due to positive assortative matching, with consequences on future earnings dynamics.

2.3.2 Workers Sorting

Now, we turn our attention to shedding light on the main drivers behind the formal/informal wage gap. As highlighted by Ulyssea (2020b), most of these studies argue that the labor market is segmented. While we do not attempt to estimate the actual wage gap between formal and informal workers, we exploit the fact that formal employees, on average, receive a higher wage to argue that workers sort into the formal and informal sector à la Roy based on their productivity. More explicitly, we use our panel data structure to demonstrate that the estimated formal premium decreases as we control for observable characteristics and individual fixed effects. Consider the following specification:

$$\log w_{it} = \beta_0 + \beta_1 \text{Formal}_{it} + \Gamma X_{it} + \delta_i + \varepsilon_{it} \tag{2}$$

where w_{it} is the wage of individual i at time t, Formalit is a dummy variable that takes the value of one if individual i is a formal worker at time t, Xit is a vector of individual characteristics, δ_i is a worker fixed effect, and ε_{it} is an error term. We estimate three different versions of β_1 in three steps. First, we run an OLS regression without X_{it} and δ_i. Next, we add the controls in X_{it} and estimate β_1. Finally, we run the full model specified in equation (2).

Table 1 presents the results of this exercise. The key takeaway is that a large part of the correlation between wages and the formal sector is explained by observable and unobservable characteristics of the worker. Specifically, when we control for observable characteristics, the coefficient drops from 0.47 to 0.355 for Chile and from 0.565 to 0.353 for Colombia. This represents a 24.47% drop from the original coefficient for Chile and 37.52% for Colombia. Next, when we add individual fixed effects, the coefficient drops even further to 0.211 for Chile and 0.093 for Colombia. This represents a further decrease of 30.64% and 46.02% from the original coefficient. Therefore, for Chile, 55.11% of the original correlation is explained by characteristics of the workers, while in Colombia, it explains a stark 83.02%. Hence, we interpret this drop in the coefficient as evidence of workers sorting into the formal and informal sectors.

One caveat of our empirical exercise is that we cannot match workers to the firms where they are employed. Hence, we cannot distinguish how much of the premium is driven by the fact that informal workers are employed by different firms than formal workers. In other words, the sorting could occur both at the worker and firm level, and the observed premium is a combination of both. While we do not take a stance on the role of firms, we acknowledge that equation (2) could still suffer from endogeneity, and therefore, we do not interpret our point estimates as causal. In fact, after controlling for both observable characteristics and individual fixed effects, our specification estimates a formal premium of 23.36% for Chile and 9.74% for Colombia. This premium can still be driven by firms and is not necessarily associated with the formal/informal sector. This is consistent with the findings by Ulyssea (2018), where the premium disappears when controlling for firm fixed effects. However, as noted in Samaniego De La Parra & Fernández Bujanda (2020), 44% of informal workers in Mexico are employed by a formal firm that also hires formal workers. This highlights the importance of studying not only sorting through the lens of firms but also from the workers' perspective.

3 Endogenous Growth Model with Informal Labor

This section develops a heterogeneous-agent endogenous growth model featuring an informal labor market. Workers exhibit productivity heterogeneity and decide whether to operate in the formal or informal sector. Operation costs vary between the two sectors, with formal workers facing labor taxes and a fixed production cost. In contrast, informal workers can avoid these costs but are subject to government fines. Additionally, workers accumulate human capital over their life cycle through interactions with others, with differing learning opportunities in each sector.

3.1 Workers Life Cycle and Production

Consider an economy where time is continuous, and there is a constant unit mass of agents. Each agent is characterized by a productivity draw coming from a continuous distribution with a cumulative probability distribution

$$G(z,t) = \Pr\left(y \leq z, \text{ at time } t\right).$$

We assume that the support of $G(z,t)$ is $\Omega = [0,\infty)$ and fixed over time. Each agent has one unit of labor and can decide to produce in the formal or informal sectors. Production technology is the same for both sectors and with productivity z, they produce z units of the final good. However, the costs an agent faces in each sector are different. Any agent working in the formal sector is subject to pay a fixed cost $F(t)$ and a proportional tax on production τ. $F(t)$ can be interpreted as the registration cost that formal businesses must incur to produce. The static return of being a formal worker is

$$Y_f(z,t) = (1-\tau)z - F(t). \tag{3}$$

In contrast, informal workers can avoid both taxes and the fixed cost. Instead, they have to pay a cost $\varphi(z,t)$ that scales with production. Intuitively, informal workers exert effort to hide their activity from the government. Similarly, $\varphi(z,t)$ can also be interpreted as the expected fine for operating informally as in Ulyssea (2018). As its production scales, the effort required to hide operations increases, or the probability of being detected by the government also increases. We assume that $\varphi(\cdot,t)$ satisfies the following properties:

$$\varphi(z,t) \geq 0, \qquad \varphi_z(z,t) \geq 0, \qquad \varphi_{zz}(z,t) > 0, \qquad \varphi(0,t) = 0,$$

where $\varphi_z(z,t)$ and $\varphi_{zz}(z,t)$ denote the first and second partial derivatives with respect to z, respectively.

The static return to being an informal worker is

$$Y_i(z,t) = \max_{y\in[0,z]} \left\{ y - \varphi(y,t) \right\}.$$

The formal worker's registering cost $F(t)$, and the expected costs paid by informal workers to hide their operation are deadweight losses. However, formal workers' labor taxes are rebated as non-distortionary lump sum transfer $T(t)$ to all workers in the economy. This modeling assumption aligns with different social welfare structures in developing economies where labor taxes finance social programs. Therefore, the only purpose of τ in this environment is to redistribute income among agents with heterogeneous skills.

Denote by $\Omega_i(t) \subseteq \Omega$ the set of workers operating in the informal sector at time t. Analogously, let $\Omega_f(t)$ be the set of workers who choose to be in the formal sector at time t. The total production of the economy is

$$Y(t) = \int_{\Omega_i(t)} Y_i(z,t)g(z,t)\mathrm{d}z + \int_{\Omega_f(t)} Y_f(z,t)g(z,t)\mathrm{d}z. \tag{4}$$

Finally, workers discount the future at a rate ρ and exit from the labor market at an exogenous

Poisson rate δ. Whenever a worker retires from the labor market, she is mechanically replaced by another worker, who takes his productivity from the current cross-sectional distribution of productivity $G(z,t)$. This implies a stationary cross-sectional age distribution $H(a)$ with a cumulative probability distribution: $H(a) = 1 - e^{-\delta a}$.

3.2 Learning

Workers can increase their productivity throughout their life cycle by interacting with others. Meetings occur at a Poisson rate α. When a worker with productivity $z(t)$ meets another worker of productivity $\tilde{z}(t)$ on a time interval Δt, they learn with the following technology:[9]

$$z(t + \Delta t) = \max\left\{z(t), \tilde{z}(t)\right\}. \tag{5}$$

As in Akcigit et al. (2018), the meetings in our economy are not symmetric in the sense that z_t can learn from $\tilde{z}(t)$, but $\tilde{z}(t)$ cannot learn from $z(t)$. In fact, $\tilde{z}(t)$ may not be looking for meetings. Conditional on having a meeting, a worker in the formal sector will meet with another formal worker with probability

$$\mathbb{P}_f^f(t) = \frac{\pi_f \mu_f(t)}{\pi_f \mu_f(t) + (1-\pi_f)\mu_i(t)}, \tag{6}$$

where $\mu_i(t)$ and $\mu_f(t)$ denote theeconomy's share of informal and formal workery. For notational convenience, we denote $\mathbb{P}_f^i(t) = 1 - \mathbb{P}_f^f(t)$. More precisely, these shares are defined as

$$\mu_i(t) = \int_{\Omega_i(t)} g(z,t)\mathrm{d}z, \qquad \mu_f(t) = \int_{\Omega_f(t)} g(z,t)\mathrm{d}z,$$

and $\mu_i(t) + \mu_f(t) = 1$ at every moment of time t. Similarly, conditional on a meeting, an informal worker will meet another with probability:

$$\mathbb{P}_i^i(t) = \frac{\pi_i \mu_i(t)}{(1-\pi_i)\mu_f(t) + \pi_i \mu_i(t)}. \tag{7}$$

The parameters π_i and π_f govern how much learning occurs within each sector. In the spirit of Jarosch et al. (2021), we allow the probability of meeting someone within your sector to differ between sectors. Furthermore, both equations (6) and (7) account for the scarcity of workers in one sector or the other. Furthermore, if we set $\pi_i = \pi_f = 1/2$, then the meetings are random and independent of the sector as in Lucas & Moll (2014). Learning opportunities will differ by sector whenever $(1-\pi_i) \neq \pi_f$. Workers then decide to sort into each sector based not only on their static return but also on their learning opportunities.

[9] This learning technology is the one used in the baseline exercises in Lucas & Moll (2014).

3.3 Sorting

Next, we consider the sorting of workers into each sector. We assume that workers can switch sectors at any time at no cost. Thus, if $V_f(z,t)$ and $V_i(z,t)$ are the value of being in the formal and informal sector, respectively, the value of a worker with productivity z is

$$V(z,t) = \max\{V_i(z,t), V_f(z,t)\}\,. \tag{8}$$

Furthermore, we can write the value of being a formal worker in a recursive form. More explicitly, letting $T(t)$ being the government's lump-sum transfer, the Hamilton-Jacobi-Bellman (HJB) equation for a formal worker of productivity z at time t is

$$\begin{aligned}(\rho+\delta)V_f(z,t) - \dot{V}_f(z,t) = {} & Y_f(z,t) + T(t) \\ & + \alpha\mathbb{P}_f^f(t)\int_{\Omega_f(t)} \max\left\{V_f(\tilde{z},t) - V_f(z,t), 0\right\}\frac{g(\tilde{z},t)}{\mu_f(t)}d\tilde{z} \\ & + \alpha\mathbb{P}_f^i(t)\int_{\Omega_i(t)} \max\left\{V_i(\tilde{z},t) - V_f(z,t), 0\right\}\frac{g(\tilde{z},t)}{\mu_i(t)}d\tilde{z}.\end{aligned} \tag{9}$$

The left-hand side (LHS) of the first line in (9) captures the net discounted value of being a formal worker, whereas the right-hand side (RHS) reflects the formal static returns. The second line in (9) captures the formal worker's human capital improvements from interacting with other formal workers. With a probability of $\alpha\mathbb{P}_f^f(t)$, a formal worker meets another formal worker. Conditional on this meeting, a formal worker with productivity z learns only from formal workers with higher human capital. Meetings with workers with skills lower than z are discarded. Finally, the last term in this equation captures the formal worker's human capital improvements from interacting with informal workers. With a probability of $\alpha\mathbb{P}_f^i(t)$, a formal worker meets an informal worker. For a formal worker with productivity z, only meetings with informal workers with higher skills increase her human capital. [10]

Similarly, the HJB equation for an informal worker is

$$\begin{aligned}(\rho+\delta)V_i(z,t) - \dot{V}_i(z,t) = {} & Y_i(z,t) + T(t) \\ & + \alpha\mathbb{P}_i^f(t)\int_{\Omega_f(t)} \max\left\{V_f(\tilde{z},t) - V_i(z,t), 0\right\}\frac{g(\tilde{z},t)}{\mu_f(t)}\mathrm{d}\tilde{z} \\ & + \alpha\mathbb{P}_i^i(t)\int_{\Omega_i(t)} \max\left\{V_i(\tilde{z},t) - V_i(z,t), 0\right\}\frac{g(\tilde{z},t)}{\mu_i(t)}\mathrm{d}\tilde{z}.\end{aligned} \tag{10}$$

The structure of the informal worker's value function, (10), is similar to the formal worker's

[10] As shown later in (11), worker's sorting implies that the last term in (9) is zero. That is, in equilibrium, formal workers do not learn from informal workers.

value function, (9). The first line's LHS captures the net discounted value of being an informal worker, while the RHS reflects the informal static returns. The second line illustrates human capital improvements by interacting with formal workers, while the last line displays human capital accumulation by interacting with other informal workers. When meeting with formal and other informal workers, an informal worker with skill z only learns from workers with higher skills.

We fully derive the HJB equations in Appendix A-1. One key feature of our model is that formal workers have increasing returns to scale on their productivity after tax and fixed costs. On the contrary, the convexity of the cost of informal implies decreasing returns to scale. This would generate a unique cutoff in a static setting that creates perfect sorting. Workers with a productivity draw above the static cutoff will sort into the formal sector, and the remaining will sort into the informal sector. This result holds also in our dynamic setting, as shown in Proposition 1.

Proposition 1. *There is exist a unique $\bar{z}(t) \in [0, \infty)$ such that*

1. $\Omega_i(t) = [0, \bar{z}(t)]$
2. $\Omega_f(t) = [\bar{z}(t), \infty)$

Moreover, $\mu_f(t) = 1 - G(\bar{z}(t), t)$ and $\mu_i(t) = G(\bar{z}(t), t)$.

Proof. See Appendix A-1. □

Observe that Proposition 1 holds if and only if

$$\forall z > \bar{z}(t), \quad V_i(z,t) < V_f(z,t),$$

and vice-versa. Moreover, at $\bar{z}(t)$, workers are indifferent between operating in the formal and informal sectors. An immediate consequence of sorting based on skills is that formal workers will only learn from more productive formal workers. On the other hand, informal workers will learn from more productive informal workers and any formal worker they meet. Formally, in equilibrium, the HJB equation (9) takes the form

$$\begin{aligned}(\rho+\delta)V_f(z,t) - \dot{V}_f(z,t) = {} & Y_f(z,t) + T(t) \\ & + \alpha \mathbb{P}_f^f(t) \int_z^\infty \left(V_f(\tilde{z},t) - V_f(z,t)\right) \frac{g(\tilde{z},t)}{1 - G(\bar{z}(t),t)} d\tilde{z}.\end{aligned} \tag{11}$$

Similarly, in equilibrium, equation (10) becomes:

$$
\begin{aligned}
(\rho+\delta)V_i(z,t)-\dot{V}_i(z,t) = {} & Y_i(z,t)+T(t) \\
& +\alpha\mathbb{P}_i^f(t)\int_{\bar{z}(t)}^{\infty}\left(V_f(\tilde{z},t)-V_i(z,t)\right)\frac{g(\tilde{z},t)}{1-G\left(\bar{z}(t),t\right)}d\tilde{z} \\
& +\alpha\mathbb{P}_i^i(t)\int_{z}^{\bar{z}(t)}\left(V_i(\tilde{z},t)-V_i(z,t)\right)\frac{g(\tilde{z},t)}{G\left(\bar{z}(t),t\right)}\mathrm{d}\tilde{z}.
\end{aligned}
\tag{12}
$$

Note that in both equations, we exploited the monotonicity of the value functions and the sorting cut-off $\bar{z}(t)$ to further simplify the expressions for the learning dynamics.

3.4 Human Capital Dynamics

We now proceed to describe the evolution of the productivity distribution over time. At every instant t, there is a set of workers learning; thus, the distribution of productivity changes over time. We note that Proposition 1 implies that the distribution dynamics differ above and below $\bar{z}(t)$. Hence, the Kolmogorov Forward Equation (KFE) will be piece-wise defined. For notation convenience, define $\lambda_{j'}^{j}(t)$ as

$$
\lambda_{j'}^{j}(t)=\frac{\mathbb{P}_{j'}^{j}(t)}{\mu_j(t)}, \quad j,j'\in\{i,f\}
$$

A natural interpretation for $\lambda_{j'}^{j}(t)$ is the quality-adjusted meeting rate of a worker in sector j' with a worker of sector j. Perfect sorting resulting from Proposition 1 creates a trade-off between the frequency of a meeting and the quality of it. Consider the case of an informal worker with productivity $z<\bar{z}(t)$. Note that

$$
\frac{\partial\mathbb{P}_i^f(t)}{\partial\bar{z}(t)}<0,
$$

because there are fewer formal workers in the economy. However, as $\bar{z}(t)$ increases, the effect on λ_i^f will depend on the probability that an informal worker meets another worker. Proposition 2 summarizes this result

Proposition 2. *Given $\pi_i,\pi_f\in[0,1]$, the quality adjusted meeting rates are monotonic on the share of informality and satisfy*

1. $\frac{\partial\lambda_i^i(t)}{\partial G(\bar{z}(t),t)}<0$ *and* $\frac{\partial\lambda_i^f(t)}{\partial G(\bar{z}(t),t)}<0$ *if and only if* $\pi_i>1/2$.
2. $\frac{\partial\lambda_f^f(t)}{\partial G(\bar{z}(t),t)}>0$ *if and only if* $\pi_f>1/2$.

Proof. See Appendix A-1. □

The intuition behind Proposition 2 is straightforward. The sorting given by Proposition 1 implies that the share of informality is $\mu_i(t) = G(\bar{z}(t), t)$. If $\pi_i > 1/2$, it implies that informal workers are more prone to meet with other informal workers. As $G(\bar{z}(t), t)$ increases, there will be more informal workers; hence, formal workers' meetings will be more scarce. Interestingly, if $\pi_i < 1/2$, then informal workers interact more with formal workers. These interactions are guaranteed to be productive, and while there is a decrease in the extensive margin of formality, the average quality of interactions will be higher.

Now, we proceed to define the KFE. We present the full derivation of the KFE in Appendix A-1. For any $z < \bar{z}(t)$, the evolution of the cumulative distribution function satisfies the following equation:

$$\frac{\partial G(z,t)}{\partial t} = -\alpha G(z,t) \left(1 - G(z,t)\lambda_i^i(t)\right) \tag{13}$$

Differentiating with respect to z yields the KFE in terms of flows

$$\frac{\partial g(z,t)}{\partial t} = \underbrace{\alpha\lambda_i^i(t)g(z,t)G(z,t)}_{\text{Inflow}} - \underbrace{\alpha g(z,t)\left(1 - \lambda_i^i(t)G(z,t)\right)}_{\text{Outflow}} \tag{14}$$

Interpreting equation (14) is useful to understand the dynamics of the model. Begin with the inflows: Due to Proposition 1, inflows can only come from informality. In particular, $G(z,t)$ workers will learn at a rate $\alpha\lambda_i^i(t)$ from a worker with productivity z with probability $g(z,t)$. Similarly, $g(z,t)$ workers will have meetings at a rate α. Conditional on a meeting, a fraction $1 - \lambda_i^i(t)G(t,z)$ will be productive meetings from which the worker will learn. The latter term can be written in a more intuitive way as follows

$$1 - \mathbb{P}_i^i(t)\frac{G(z,t)}{G(\bar{z}(t),t)}$$

Hence, the term is the complement probability of a worker with productivity z meeting a less productive worker. Correspondingly, for any $z \geq \bar{z}(t)$, the KFE that describes the dynamics of distribution is

$$\frac{\partial G(z,t)}{\partial t} = -\alpha\left(1 - G(z,t)\right)\left[\lambda_i^f G(\bar{z}(t),t) + \lambda_f^f\left(G(z,t) - G(\bar{z}(t),t)\right)\right] \tag{15}$$

and in terms of flows

$$\frac{\partial g(z,t)}{\partial t} = \underbrace{\alpha\lambda_i^f(t)G(\bar{z}(t),t)g(z,t)}_{\text{Inflow from informality}} + \underbrace{\alpha\lambda_f^f(t)\left[G(z,t) - G(\bar{z}(t),t)\right]g(z,t)}_{\text{Inflow from formality}} - \underbrace{\alpha\lambda_f^f(t)g(z,t)\left(1 - G(z,t)\right)}_{\text{Outflow}} \tag{16}$$

Equation (16) has a similar interpretation as equation (14), but now inflows come from both formality and informality. Note that for a fixed worker, productivity is a non-decreasing function of time that discontinuously jumps. But as the distribution shifts to the right, the informality threshold $\bar{z}(t)$ also shifts to the right. Consequently, a worker who initially sorted in the formal sector and by chance never had a productive meeting will eventually be caught by the informality threshold $\bar{z}(t)$ and, therefore, sorted in the informal sector. With these, we can define an equilibrium in our economy.

Definition 1. *Given an initial distribution $g(z, 0)$, an equilibrium is a tuple of functions (V, V_i, V_f, g) from $\mathbb{R}^2_+$ to $\mathbb{R}$ and a function $\bar{z}$ from $\mathbb{R}$ to $\mathbb{R}$ such that $\forall t \geq 0$:*

1. *$V_f(z, t)$ satisfies equation (11), and $V_i(z, t)$ satisfies equation (12),*
2. *For every $z < \bar{z}(t)$, $g(z, t)$ satisfies equation (14),*
3. *For every $z \geq \bar{z}(t)$, $g(z, t)$ satisfies equation (16),*
4. *$V(z, t)$ satisfies equation (8),*
5. *The indifference condition $V_f(\bar{z}(t), t) = V_i(\bar{z}(t), t)$ is satisfied,*
6. *The government has a balanced budget, satisfying*

$$T(t) = \int_{\bar{z}(t)}^{\infty} \tau z g(z, t) \mathrm{d}z \tag{17}$$

The first condition in the definition of the equilibrium states that the value functions for formal and informal workers, $V_f(z, t)$ and $V_i(z, t)$, are consistent with workers sorting. Conditions 2 and 3 estate that human capital dynamics are dictated by the KFE equations, also consistent with worker sorting. Moreover, conditions 4 and 5 indicate that workers freely sort into the sector that yields higher returns and that there exists a marginal worker with skill $\bar{z}(t)$ who is indifferent between both sectors. Finally, the last condition states that the government runs a balanced budget, and this implies that per-capita transfers are equal to labor taxes levied on formal workers.

3.5 Balanced Growth Path

In the remainder of the paper, we focus on a particular equilibrium. In particular, as is common in the economic growth literature, we focus on a Balanced Growth Path (BGP) equilibrium. Intuitively, a BGP is an equilibrium satisfying the conditions in Definition 1 in which the growth rate of the economy and the relative human capital distribution are constant over time. Formally,

Definition 2. *A balanced growth path (BGP) is a vector $(\gamma, \bar{x})$ and a tuple of real functions $(v, v_i, v_f, \phi, \Phi)$ defined over $\mathbb{R}_+$ such that $\bar{z}(t) = \bar{x}e^{\gamma t}$ and*

$$V(z,t) = e^{\gamma t} v\left(z e^{-\gamma t}\right) \qquad V_i(z,t) = e^{\gamma t} v_i\left(z e^{-\gamma t}\right)$$
$$V_f(z,t) = e^{\gamma t} v_f\left(z e^{-\gamma t}\right) \qquad g(z,t) = e^{\gamma t} \phi\left(z e^{-\gamma t}\right)$$

for every pair (z,t) and (V, V_i, V_f, g) together with $\bar{z}(t)$ define an equilibrium with initial distribution $g(z,0) = \phi(z)$ and $G(z,0) = \Phi(z)$.

From this definition, we see that a BGP is a path for the skill distributions for formal and informal workers in which all quantiles grow at the same rate γ. To further characterize the values of being a formal or informal worker in the BGP, we specialize the functional forms for the formal fixed cost, $F(t)$, and the expected informality cost, $\varphi(z,t)$. Specifically, we assume that:

$$F(t) = F e^{\gamma t}, \quad \text{and} \quad \varphi(z,t) = \eta e^{-\gamma t} z^2, \tag{18}$$

for some constants, $F > 0$ and $\eta > 0$. Equation (18) reveals that formal workers' fixed cost is a fraction of the total production in the economy. Therefore, as the economy grows, sunk costs of operating formally also grow.[11] Similarly, the cost of informality, $\varphi(z,t)$, has two components. While the first component, $\eta e^{-\gamma t}$, decreases over time, the second component z^2 increases as the economy grows. Because in the BGP, the distribution of productivity shifts to the right, improvements in human capital make it more costly to operate informally. Hence, the first component of $\varphi(z,t)$ decreases in such a manner that offsets the increasing cost due to distributional changes.

Letting $x \equiv z e^{-\gamma t}$ be the relative human capital of a given worker, equation (17) implies that government transfers in the BGP, T, are equal to

$$T = \int_{\bar{x}}^{\infty} \tau x \phi(x) \mathrm{d}x = \tau \left(1 - \Phi(\bar{x})\right) \mathbb{E}\left[x | x \geq \bar{x}\right] \tag{19}$$

With this notation, along the BGP, the HJB equation for formal workers takes the form:

$$\begin{aligned}(\rho + \delta - \gamma) v_f(x) + \gamma x v_f'(x) = {} & (1-\tau)x - F + T \\ & + \alpha \lambda_f^f(\bar{x}) \int_x^{\infty} \left(v_f(\tilde{x}) - v_f(x)\right) \phi(\tilde{x}) d\tilde{x},\end{aligned} \tag{20}$$

where we replaced the value for $Y_f(\cdot, t)$, $F(t)$, and $T(t)$ in the BGP. Similarly, the HJB equation for an informal worker with relative human capital x is

[11] This might reflect the fact that the government has limited capacity to register firms. Therefore, as national production increases, firms need to spend more time or resources to be formally registered.

$$
\begin{aligned}
(\rho+\delta-\gamma)v_i(x)+\gamma x v_i'(x) = x - \eta x^2 + T & \\
+ \alpha\lambda_i^f(\bar{x}) \int_x^{\infty} \left(v_f(\tilde{x}) - v_i(x)\right)\phi(\tilde{x})d\tilde{x} & \\
+ \alpha\lambda_i^i(\bar{x}) \int_x^{\bar{x}} \left(v_i(\tilde{x}) - v_i(x)\right)\phi(\tilde{x})d\tilde{x} & \qquad (21)
\end{aligned}
$$

Finally, to characterize the growth rate along the BGP, more structure on the initial distribution of skills is needed. Therefore, we add an assumption on the initial productivity distribution. Although our results do not hinge on this assumption, it gives us closed-form solutions for the economy's growth rate, γ.

Assumption 1. *The initial productivity distribution, $G(z,0)$, has a Pareto tail. That is, there exist $k, \theta > 0$ such that*

$$
\lim_{z\to\infty} \frac{1-G(z,0)}{z^{-1/\theta}} = k \tag{22}
$$

Assumption 1 implies that the initial distribution tail approaches 0 at the same rate as a Pareto distribution but encompasses a higher class of possible distributions. In addition, note that the support of the initial distribution is unbounded, suggesting that at time 0, all the knowledge in the economy already exists, and it is just waiting to be discovered. Given the definition of a BGP, the distribution moves steadily to the right as time progresses. Hence, very unlikely knowledge becomes feasible. Finally, Assumption 1 is a sufficient condition for the economy to sustain a growth rate $\gamma > 0$.

Proposition 3. *Suppose that Assumption 1 holds and $\pi_i < 1$. Then there is a number γ that holds a BGP, and it's defined as*

$$
\gamma = \alpha\theta\left[\Phi(\bar{x})\lambda_i^f + (1-\Phi(\bar{x}))\lambda_f^f\right] \tag{23}
$$

Proof. See Appendix A-1 □

Equation (23) gives a clear interpretation of the growth rate along the BGP. First, growth depends positively on α, which governs the rate at which agents interact. Thus, an economy where agents meet frequently will have a higher growth rate. Second, the growth rate increases with θ. Recall that $1/\theta$ captures the dispersion of productivity in the economy. In an extreme case where $\theta \to 0$, the distribution becomes degenerate, and there is no dispersion. Hence, the growth rate increases as productivity is distributed more evenly across agents. We assume that $\theta \in [0,1]$ so that the distribution has a well-defined first moment. Finally, the latter term is an endogenous object that encapsulates the weighted average rate at which agents meet formal workers. Proposition 1 implies that more productive workers sort into the formal sector, and consequently, interactions with formal workers yield higher learning, translating into higher economic growth.

One useful case to study is the random matching case where $\pi_i = \pi_f = 1/2$. This is the same case studied by Lucas & Moll (2014) with a constant meeting rate α. Note that, for this particular case,

$$\lambda_i^i = \lambda_i^f = \lambda_f^i = \lambda_f^f = 1,$$

as all agents are equally likely to meet. In this specific scenario, sorting has no implications for learning; hence, the growth rate of the economy is

$$\gamma = \alpha\theta$$

This case highlights a critical property of our model. While sorting is static, it has dynamic implications as it changes learning opportunities. When we remove friction segmentation in terms of learning, the growth rate depends solely on the frequency of meeting rates and the dispersion of productivity across agents.

4 Model Estimation

In this section, we describe how we estimate the parameters of the model using the data from Colombia.[12] We also highlight the variation in the data that identifies these parameters.

4.1 Estimation

The model developed in Section 3 has 10 parameters:

$$\Gamma = \left\{ \underbrace{\tau, F, \eta}_{\text{Regulatory}}, \underbrace{\alpha, \pi_i, \pi_f}_{\text{Interactions}}, \underbrace{k, \theta}_{\text{Distributional}}, \underbrace{\rho, \delta}_{\text{Macro}} \right\}$$

The first three parameters are meant to capture the regulatory status of the economy: τ is the tax on formal production, F is the fixed cost of formality, and η is a parameter that governs the levels of the cost of informality.

The next set of parameters is all related to interactions. First, α is the Poisson arrival rate for a meeting with another agent. Furthermore, π_i and π_f are measures of how the economy is segregated in terms of learning. In the extreme case where $\pi_f = 1$ and $\pi_i = 1$ implies that learning occurs only within each sector. Moreover, two parameters describe the initial distribution of productivity: θ, which speaks to the dispersion, and k, which anchors the units of the distribution. Finally, ρ and δ are macro parameters that capture the discount rate of agents and the Poisson rate at which agents retire from the economy.

[12] The model estimation for Chile is still a work in progress.

Of the 10 parameters in our model, we externally calibrate two of them. First, we set $\rho = 0.05$ following Akcigit et al. (2021). Furthermore, we normalize k to 2. This parameter is not identified as it anchors the units of the productivity distribution. However, this is an unobserved object in the data, and hence we cannot identify it. Anchoring the units of the distribution for both countries allows us to do cross-country comparisons.

Next, we directly match δ from the data. Note that our implied age distribution $H(\cdot)$ yields an equation that allows us to estimate δ from the data. Explicitly note that

$$\log\left(1 - H(a)\right) = -\delta a.$$

Hence, we run a regression of age a on the empirical cumulative distribution function, $H^e(a)$ without a constant.

We estimate the remaining seven parameters, $\Gamma' = \{\tau, F, \eta, \alpha, \pi_i, \pi_f, \theta\}$ by employing a Simulated Method of Moments (SMM) estimator. Formally, let $M(\Gamma')$ be the vector of size S of moments generated by the model with parameters Γ'. Analogously, let M^e be the vector of empirical moments from the data. We estimate our model by minimizing the distance between the empirical and the simulated moments. Given a vector of empirical moments M^e, our score function is

$$\Psi\left(M^e\right) = \min_{\Gamma'} \sqrt{\sum_{m=1}^{S} \left[\frac{\omega_m\left(M_m^e - M_m(\Gamma')\right)}{{}^1\!/\!{}_2|M_m^e| + {}^1\!/\!{}_2|M_m(\Gamma')|}\right]^2}$$

where ω_m is the relative weight of moment m. To find a global minimum, we implement the TikTak algorithm described by Arnoud et al. (2019) to search over the parameter space. In every iteration of the algorithm, we solve the equilibrium and calculate the simulated moments to compute the score function. We next proceed to discuss how our parameters are identified.

4.2 Identification

Now, we turn our attention to describing how our selected moments identify each of our parameters. We select moments associated with informality, growth, and the transition probability between sectors, as well as wage growth. While we employ an indirect inference estimation strategy, and all our parameters are identified jointly, we provide a heuristic discussion on how the selected moments aid in identifying our parameters.

Parameters determining dynamic returns. There are four parameters governing the dynamic returns for workers: α governs the frequency of meetings, θ governs the dispersion of the productivity distribution, influencing the returns of learning. Additionally, π_i and π_f represent the probability of interacting with other workers from the same sector, conditional on a meeting. To identify the dynamic parameters, we estimate the transition matrix between the formal and informal sectors. To achieve this, we exploit the panel structure of our data and construct the empirical transition matrix based on workers who switch sectors within each wave of the survey.

However, since the frequency of our survey is not yearly, we first need to calculate the annual transition matrix to match the model. Let g be the number of years within each wave of the survey, and P^e be the empirical transition matrix. To calculate the 1-year transition, we diagonalize P^e and then raise the diagonal to the power of the inverse of the gap between waves, denoted as g. More formally, let D be a diagonal matrix such that

$$P^e = A^{-1} D A,$$

for some matrix A. Then, the 1-year transition matrix can be expressed as

$$\tilde{P}^e = A^{-1} D^{1/g} A.$$

By definition, $\tilde{P}^e$ is a 2×2 matrix which satisfies

$$\sum_{i,j \in \{f,i\}} \tilde{P}^e_{ij} = 1,$$

determining two different moments. We target $\tilde{P}^e_{i,f}$ and $\tilde{P}^e_{f,i}$. Note that our model yields a closed-form solution for both. The probability of an informal worker transitioning into a formal worker in a year is

$$\Pr(i \to f) = \alpha(1 - \pi_i).$$

Alternatively, the probability of a formal worker can be written as

$$\Pr(f \to i) = \int_{\bar{x}}^{\bar{x}e^{\gamma}} \alpha \left[1 - \pi_f \left(1 + \Phi(\bar{x}) - \Phi(x)\right)\right] \frac{\phi(x)}{1 - \Phi(\bar{x})} \mathrm{d}x$$

We provide a more formal derivation in Appendix A-1.1.

Equation (23) provides an equation that relates GDP growth to all four of our dynamic parameters. Additionally, we target the share of informal workers in the economy. In our model, the share of informality is expressed as $\Phi(\bar{x})$, aiding us in calculating both the growth rate and the transition probability from formal to informal.

Parameters determining static returns: There are three parameters that govern the static returns of workers: τ represents the proportional tax on productivity, F stands for the fixed cost for formal workers, and η governs the cost of informality. Equation (8) implies that sorting considers both static returns and learning opportunities. To identify the static parameters, we focus on wage growth over the life cycle. Specifically, we target each point in Figure 2.

We delve deeper into the details of how each parameter identifies each moment by presenting a

Table 2: Estimated Parameters - Colombia

Parameter	Description	Value
k	Pareto Scale Parameter	2
ρ	Discount Rate	0.05
δ	Death Hazard Rate	0.045
F	Fixed Cost	0.201
θ	Tail of Productivity Distribution	0.371
π_f	Probability of Meeting within Sector (Formals)	0.855
π_i	Probability of Meeting within Sector (Informals)	0.28
α	Meeting Rate	0.081
τ	Formality Tax	0.204
η	Informality Cost	3.344

Notes: Table 2 presents the Simulated Method of Moments estimation results for Colombia.

sensitivity matrix for each moment in Appendix A-1.

4.3 Estimation Results

Table 2 presents the estimation results for Colombia. There are several results worthy of highlighting. First, interactions are rare in our economy. The estimated value of α implies that the probability of a worker interacting with another worker within a year is 7.1%. This result aligns with the findings of Lagakos et al. (2018), which indicate that developing economies generally experience a less pronounced wage increase over the life cycle. Second, the estimate of the tail of the initial productivity distribution, θ, is close to the one estimated by Lucas Jr (2009), $\theta = 0.5$, who used a similar model to match the variance of earnings in the US. Our estimate of $\theta = 0.371$ reflects the lower variance of wages for workers at the end of their careers in Colombia relative to the US. Third, the estimated labor formal tax τ closely resembles the redistributive labor taxes in the Colombian regulation: 20.5% (12% for pension funds and 8.5% for healthcare). While we did not directly utilize these values to estimate τ, our estimate, targeting the moments discussed in the previous section, serves as an over-identification exercise. In doing so, we can accurately replicate the actual Colombian labor tax regulations. Fourth, the estimated informality cost η lies closely to the one estimated by Ulyssea (2018) using a similar functional form to (18) and using data from Brasil, $\eta \approx 5.01$.

The estimation results also illustrate insights into the learning frictions between formal and informal workers. Conditional on meeting another worker, formal workers meet other formal workers with a probability of 85%. Similarly, conditional on a meeting, informal workers interact with formal workers with a probability of 70%. This suggests that there are asymmetries in learning environments in the formal and informal sectors. While formal workers tend to always interact within the formal sector, informal workers also tend to interact with formal workers. These results suggest that barriers for informal workers to interact with their informal counterparts are quantitatively small and are in line with the no-dual economy results described by Ulyssea (2020a). Based on Proposition 2, this implies that in our estimated economy, both λ_i^f and λ_f^f are monotonically

Table 3: Goodness of Fit - Colombia

	Model	Data	Weight
Growth Rate (%)	3.361	3.234	1/6
Informality Rate (%)	63.455	63.100	1/6
Wage Growth (Informal), 25-34	0.012	0.174	1/24
Wage Growth (Informal), 35-44	0.013	0.241	1/24
Wage Growth (Informal), 45-54	0.022	0.253	1/24
Wage Growth (Informal), 55-64	0.025	0.179	1/24
Wage Growth (Formal), 25-34	0.075	0.236	1/24
Wage Growth (Formal), 35-44	0.080	0.364	1/24
Wage Growth (Formal), 45-54	0.129	0.429	1/24
Wage Growth (Formal), 55-64	0.193	0.453	1/24
Transition Probability - I-F	0.056	0.060	1/6
Transition Probability - F-I	0.074	0.065	1/6

Notes: Table 3 presents the Simulated Method of Moments estimation goodness of fit and moment weights.

increasing on the share of informality $\Phi(\bar{x})$.

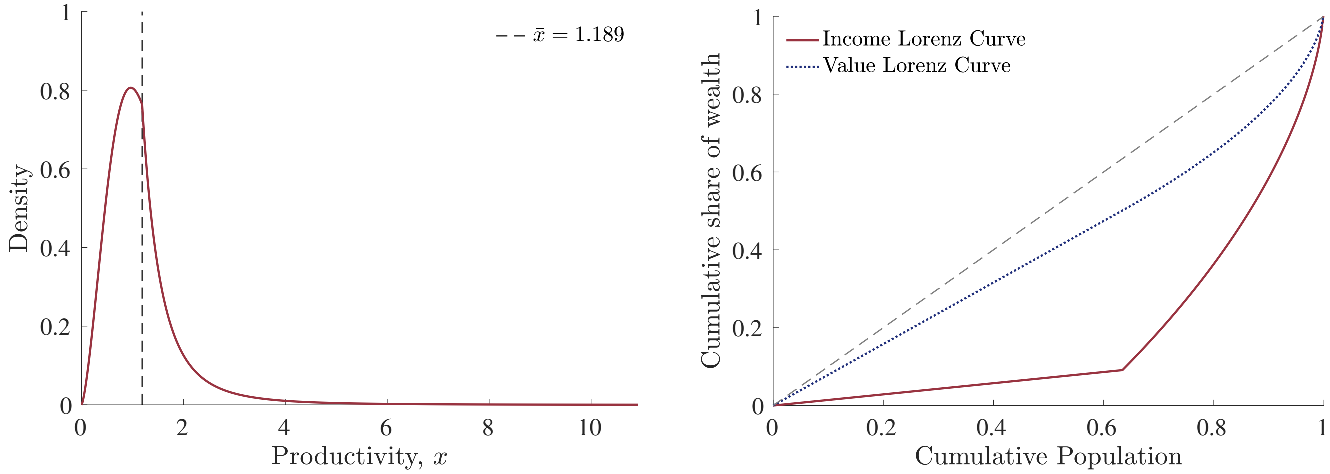
Table 3 presents the goodness of fit of the estimated model. The model closely matches the aggregate growth rate, the informality rate, and the yearly transition rates between formal and informal sectors. As displayed in the third column, these are the moments with a larger weight in our SMM estimation procedure. In this sense, the estimated model closely replicates key economic aggregates of the Colombian economy. Nevertheless, as discussed below, we still need to improve the model's ability to replicate the earnings paths for individual workers.

Table 3 also illustrates that the estimated model falls short of matching the wage growth paths for both formal and informal workers. This result is most likely to be implied by the low value of our estimate for α. We believe that there are potentially two features of the model that are causing these low-wage growth trajectories. First, there is the assumption that new entrants to the labor market draw their productivity from the stationary distribution. In this sense, workers who enter the labor market later in time benefit from human capital improvements from previous generations. Fixing the entry distribution of skills different from the stationary one could allow the model to generate steeper wage growth paths as new entrants do not have this later mover advantage, accumulating more human capital over the life cycle to catch up with incumbent workers.

Alternative learning technologies could also provide wage growth paths closer to the ones observed in the data. Learning technology (5) does not have barriers to learning. This means that the least skilled worker can acquire the skill of the most skilled worker in the economy, conditional on a meeting. Hence, to rationalize the patterns in the data, the model underestimates the meeting probability, α.[13] In reality, it is plausible that workers only acquire a fraction of the higher skills of the counterparts they meet. Similarly, workers could benefit relatively more by interacting with

[13] Given the baseline learning technology, larger values of α are likely to deliver unrealistically steeper wage growth paths.

Figure 3: Estimated Equilibrium



(a) Productivity Distribution

(b) Lorenz Curves

workers with similar skills. Thus, in the worker in progress, we are exploring alternative learning technologies such as those introduced by Lucas & Moll (2014), which incorporate features such as symmetric meetings, limits to learning, and exogenous learning shocks.

To conclude this section, Figure 3 displays different features of the BGP equilibrium with the estimated parameters. First, Figure 3(a) illustrates the stationary density of the relative productivity, x, and the estimated formality cutoff, $\bar{x} = 1.189$. The stationary human capital distribution, $\phi(x)$, inherits some of the properties of the initial Pareto skill distribution, $g(z, 0)$. In particular, the distribution exhibits a long tail that encodes a large dispersion for higher skill levels. Interestingly, in the BGP, the human capital stock of the most skilled formal workers (percentile 95th) is 2.14 times larger than the level of human capital of the least skilled formal worker (marginal worker with $x = \bar{x}$). In contrast, human capital dispersion in the informal sector is significantly larger, with the human capital level of the most skilled informal worker being 3.6 times larger than the level of human capital of the least informal worker (percentile 5th). This result is driven by the larger share of informal workers in the economy as a whole.[14]

Figure 3(b) plots two Lorenz curves for the estimated BGP equilibrium. The solid (red) line shows the fraction of the total current income attributed to workers with productivity less than x. The cumulative share of wealth (measured as current income) is linear until the cumulative population reaches 63%, which is the formality cutoff. From this point onwards, the cumulative share becomes a concave function, reflecting the large skill dispersion of the invariant human capital distribution. The dashed (blue) line represents the fraction of the total present discounted income attributed to workers with productivity less than x. In contrast to the current income curve, the Lorenz value curve does not have any kinks, illustrating the smooth pasting property of the formal/informal value functions (20) - (21).

The income Lorenz curve exhibits less inequality than the income Lorenz curve. In line with the

[14]Note that the formality cutoff $\bar{x}$ is at the percentile 63th of the human capital invariant distribution.

results in Jarosch et al. (2021), this property highlights the importance of examining present value rather than flow Lorenz curves in dynamic problems. Importantly, the value Lorenz curves account for the learning opportunities for both formal and informal workers. Interestingly, accounting for dynamic effects increases the reduction in inequality relatively more for low-skilled workers.

5 The Role of Formalization Policies

In this final section, we present two counterfactual exercises that quantify the aggregate effects of implementing two types of formalization policies. These policies are aimed at reducing the size of the informal sector and increasing government revenue. Moreover, they have become increasingly relevant, as governments invest a large amount of resources in implementing them. However, such policies can come in the form of sticks or carrots. We analyze the effects of policies that aim to prosecute informal operations separately from regulations seeking to incentivize the formalization of informal businesses. With our estimated parameters, we are able to study a counterfactual economy where the costs of formality and informality are different. In particular, we consider the effects of policies that are intended to increase the formality rate in the economy. In terms of our model, we investigate how the economy reacts to an increase in the cost of informality, η, versus a decrease in the fixed cost of formality, F. More explicitly, we explore the effects on growth, informality, government revenue, and welfare. In particular, welfare is defined as the mean-adjusted value across all workers with different human capital levels:

$$W^* = \frac{\mathbb{E}_\Phi \left[v(x) \right]}{\rho + \delta - \gamma}. \tag{24}$$

We begin by computing the counterfactual economy upon a change in the cost of informality, and then we proceed to do the same exercise, changing the formal sector fixed cost.

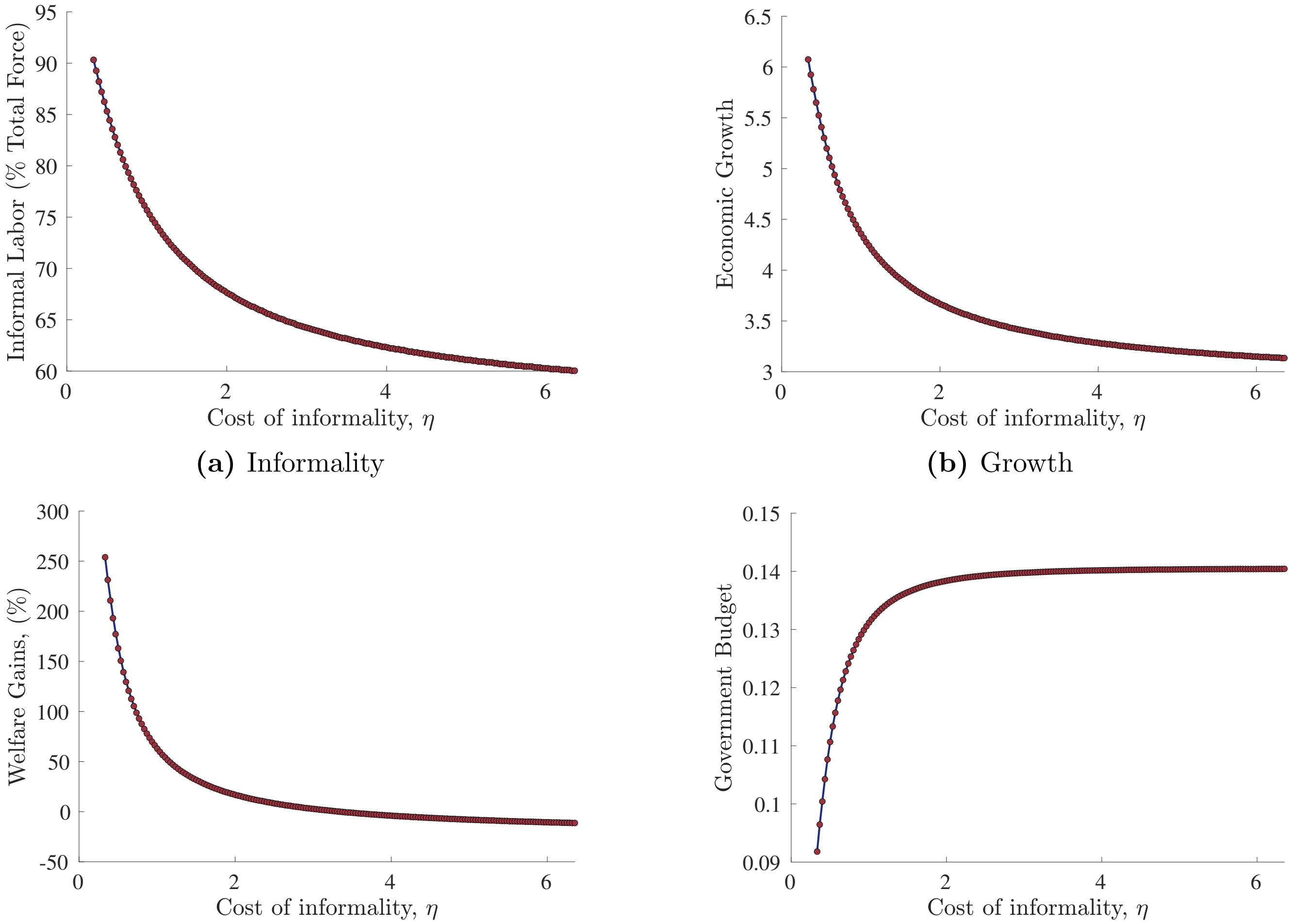
5.1 The Cost of Informality

Prosecuting informal activity is something local authorities exert a lot of effort and resources. For example, in 2023, Colombia's tax authorities increased auditing efforts and closed 355 establishments in the country (El Tiempo, 2023).

We model an increase in the cost of informality by changing the parameters that govern the informality cost $\varphi(z,t)$. In particular, given our assumed functional form in (18), we model a change in government prosecution of informality as a change to η. More explicitly, we take our estimated parameter from Table 2 and create a grid around that point. We set the lower bound of the grid as 50% of the estimated parameters and the upper bound as a 50% increase. Then, for each point on the grid, we solve the equilibrium and calculate the moments of interest. Figure 4 illustrates the results of this exercise.

In the first place, Figure 4(a) shows that the share of informality is decreasing with η. As expected, higher values of η imply a higher cost of operating in the informal sector, and therefore, fewer workers would be willing to sort into it. More interestingly, the exercise shows that the share of

Figure 4: Increase in the cost of informality, η



(a) Informality

(b) Growth

(c) Economic Welfare

(d) Government Budget

Notes: All Figures depict the effect of changing the magnitude of the constant term from the cost of informality η while keeping the rest of the parameters constant. We solve the model for every value of η in a grid and plot the resulting values. Figure 4(a) depicts the percent of workers sorting into the informal sector. Figure 4(b) presents the growth rate of the economy. Figure 4(c) shows the welfare changes defined in (24) (consumption equivalent). Finally, Figure 4(d) illustrates the government budget as defined in (17).

informality has a convex shape in η, meaning that there are marginal diminishing returns from increasing the cost of informality. Therefore, these policies are very effective when the informal sector is too large, but as it decreases, they start to become less effective.

Interestingly, Figure 4(b) shows an inverse relationship between η and economic growth, γ. As informality increases, economic growth dampens. Although this might seem like a striking fact at first, it becomes more intuitive when considering the source of growth in our model. In particular, our growth rate hinges on the interaction with formal workers. Therefore, with an increase in the cost of informality, the marginal worker who decides to formalize is less productive than all existing formal workers. Consequently, as more workers formalize, the average quality of interactions with formal workers decreases. Furthermore, the number of interactions for formal workers also decreases as the recently formalized workers crowd out the learning opportunities. Note that during this exercise, we fixed both α and θ, which are primitives of the model that directly impact growth. This exercise shows the importance of sorting in our model. While we only change the static incentives to sort into sectors, the dynamic implications of such policies are major.

Figure 4(c) shows that economic welfare is decreasing with the value of η. On the one hand, higher η increases government revenue and hence the levels of the value function for low-skilled workers. On the other hand, welfare depends positively on growth, and therefore, an increase in η has a negative effect on welfare. Although static returns increase at every period, the growth rate is sufficiently reduced so that, in aggregate, the net effect is negative. Nonetheless, since there are two opposing forces at play, the degree of convexity is lower than in Figure 4(b). Moreover, as η increases, it is more costly to produce for those workers that remain operating informally, having a first-order negative effect on the value for the least skilled workers in the economy.[15]

Finally, Figure 4(d) illustrates that there is an increasing relationship between the government budget and the cost of informality. Indeed, as the share of informality decreases, more workers are paying taxes and fixed costs. Consequently, as the government rebates its income as lump-sum transfers, the level of the value function increases. Equation (24) shows that this will mechanically have a positive effect on economic welfare as consumption increases for all workers in the economy.

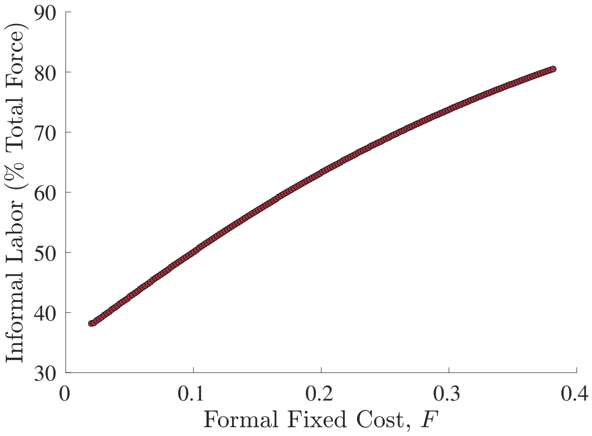
5.2 The Cost of Formalization

As discussed by Ulyssea (2020a), other types of policies that could decrease informality are the ones aimed at reducing the costs of operating formally. In our framework, two parameters govern the costs for formal workers: the payroll tax, τ, and the fixed cost of production, F. As payroll taxes are aimed to fulfill a redistributive purpose in many developing countries, we focus on the effects of changing the fixed cost F. This could be interpreted as governments losing up on red tape costs to register formal production. Similarly to our previous counterfactual exercise, we take our estimated value for F and create a grid around that point, setting the lower and upper bounds of the grid as 50% decrease and increase of the estimated parameter, respectively. Solving for the BGP in each point on the grid, Figure 5 displays the results.

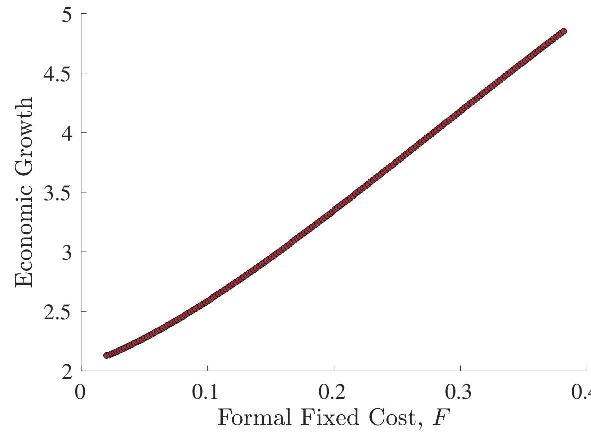
Decreasing formal workers' registering costs has similar qualitative effects as increasing the cost of informality. Nevertheless, there are some quantitative differences worth highlighting. First, Figure 5(a) illustrates that reducing formal sector registration costs by half reduces the share of informal workers from 63% to 48%. Compared to the results in 4(a), reducing formality costs is much more effective at reducing the size of the informal labor markets than increasing the cost of informality. Furthermore, Figure 5(b) illustrates again the trade-off between the size of the informal sector and economic growth. A decrease in F is effective in formalizing marginally unproductive workers who crowd out the interactions in the formal sector. The negative long-run effects of formalization on economic growth translate into negative welfare gains, as is shown in Figure 5(c). Finally, a lower growth rate also implies less government revenue, as illustrated in Figure 5(d).

[15] Figure 4(c) shows that they are large welfare gains of reducing η. This is mainly driven by the gains in production by the least skilled workers: as η decreases, it is easier to produce informally, and therefore welfare increases.

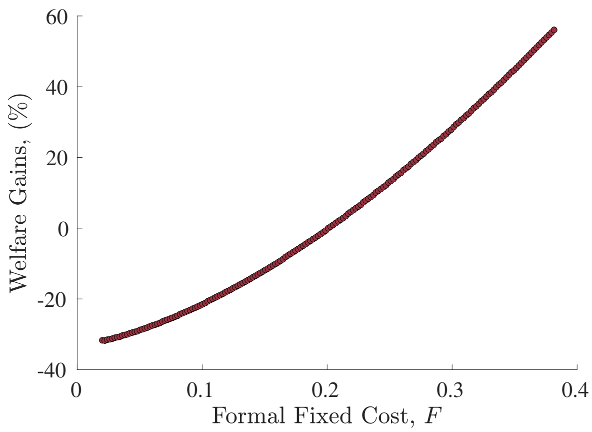
Figure 5: Decrease of formal sector fixed cost, F



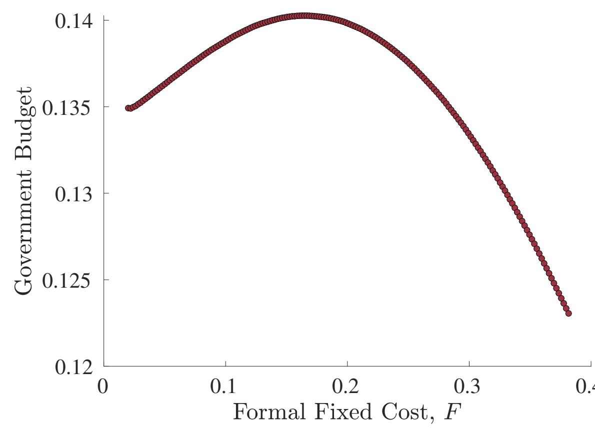
(a) Informality



(b) Growth



(c) Economic Welfare



(d) Government Budget

Notes: All Figures depict the effect of changing the magnitude of the constant term from the formal fix cost F while keeping the rest of the parameters constant. We solve the model for every value of η in a grid and plot the resulting values. Figure 5(a) depicts the percent of workers sorting into the informal sector. Figure 5(b) presents the growth rate of the economy. Figure 5(c) shows the welfare changes defined in (24) (consumption equivalent). Finally, Figure 5(d) illustrates the government budget as defined in (17).

5.3 Discussion

The aggregate effects of the two formalization policies are somewhat puzzling. According to our estimates in Table 2, reducing the size of the informal sector has adverse effects on both growth and welfare. This outcome results from the interplay of two forces. First, as the formal sector expands, more workers gain access to better learning opportunities. Since $\pi_{ff} > \pi_{if}$, the probability of encountering a high-skill worker, given a meeting, is higher in the formal sector. Therefore, formalizing workers positively impacts long-term growth, as expected gains in human capital increase when meeting higher-skilled workers. However, the formalization of workers also exerts a negative effect on the quality of interactions within the formal sector. As the skills of the marginal workers who formalize are lower, formal workers start interacting with counterparts of lower skill levels, hindering improvements in human capital. Based on our current quantification, the negative effects of formalization policies outweigh the positive ones.

The effects of the formalization policies shed light on within-group learning dynamics. The results in Figures 4 and 5 suggest potential long-term gains from segmenting the formal and informal markets when learning opportunities differ across both sectors. Specifically, these figures indicate that reducing the size of the formal sector could be beneficial in the aggregate. The learning technology in equation (18) might be the key feature driving this result. Since anyone can learn from the most skilled workers, the economy benefits from a smaller formal sector. With fewer formal workers, the quality of interactions within the sector improves, effectively increasing the mass of workers at the far right tail of the human capital distribution. Given our estimate of $\pi_{if} \approx 0.7$, informal workers frequently interact with formal workers. Consequently, they benefit from advances 'at the frontier' of the skill distribution, leading to an increase in expected human capital improvements upon meeting a formal worker. In summary, the economy thrives in a scenario where the best workers initiate learning among themselves, expanding the knowledge frontier, and subsequently, the rest of the workers catch up.

6 Conclusion

This study sheds light on the complex dynamics of informal labor markets in developing economies, specifically in the context of Chile and Colombia. By providing empirical evidence on wage differentials and growth patterns for formal and informal workers, we offer insights into the intricate relationship between human capital formation, labor market segmentation, and economic outcomes. The observed wage disparities, influenced by workers' sorting based on skill and divergent human capital accumulation over the life cycle, underscore the importance of understanding the mechanisms that drive the formal-informal wage gap. Moreover, our developed heterogeneous agent endogenous growth model offers a theoretical framework to rationalize these empirical findings, emphasizing the role of learning opportunities, fixed and variable costs, and the interplay between the formal and informal sectors in shaping economic outcomes. In the policy realm, our counterfactual analyses underscore the nuanced impacts of formalization policies. While efforts to reduce informality by either increasing the cost of informal operations or decreasing formalization costs may seem intuitive, our results caution against overlooking the intricate dynamics at play. The crowding-out effect on human capital in the formal sector, stemming from the movement of less-skilled workers, challenges the conventional wisdom regarding the benefits of formalization.

In our work in progress, we are expanding the model along two dimensions. First, we are exploring alternative learning technologies that better capture the dynamics reflected in the wage paths observed in the data. We believe that by considering these different technologies, the quantitative effects of formalization policies will become more plausible. Specifically, we suspect that technologies in which workers can only appropriate a fraction of the skills of more knowledgeable workers will lead formalization policies to have positive impacts on growth and welfare. Second, our current framework attributes all wage improvements to workers' human capital accumulation. In reality, it is plausible that firm productivity improvements over time also influence the observed wage paths. Therefore, we are working to incorporate technological change from the firm side into our framework. Assessing the relative importance of workers' human capital accumulation and firm productivity improvements on growth and welfare, and understanding how worker sorting between formal and informal sectors affects both channels, is crucial for effective policy design.

References

Akcigit, U., Akgunduz, E., Alp, H., Cilasun, S. M., & Quintero, J. M. (2022, October). *Cost of Size-dependent Regulations: The Role of Informality and Firm Heterogeneity* [Unpublished Manuscript].

Akcigit, U., Alp, H., & Peters, M. (2021, January). Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries. *American Economic Review*, *111*(1), 231–275. Retrieved 2022-11-08, from http://www.aeaweb.org/articles?id=10.1257/aer.20180555 doi: 10.1257/aer.20180555

Akcigit, U., Caicedo, S., Miguelez, E., Stantcheva, S., & Sterzi, V. (2018, March). *Dancing with the Stars: Innovation Through Interactions.* Cambridge, MA: National Bureau of Economic Research. Retrieved 2023-02-02, from http://www.nber.org/papers/w24466.pdf doi: 10.3386/w24466

Arnoud, A., Guvenen, F., & Kleineberg, T. (2019, October). *Benchmarking Global Optimizers* [Working Paper]. National Bureau of Economic Research. Retrieved 2023-12-22, from https://www.nber.org/papers/w26340 doi: 10.3386/w26340

Bargain, O., & Kwenda, P. (2014, October). The Informal Sector Wage Gap: New Evidence Using Quantile Estimations on Panel Data. *Economic Development and Cultural Change*, *63*(1), 117–153. Retrieved 2023-10-23, from https://www.journals.uchicago.edu/doi/10.1086/677908 doi: 10.1086/677908

Ben-Porath, Y. (1967, August). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*, *75*(4, Part 1), 352–365. Retrieved 2024-01-05, from https://www.journals.uchicago.edu/doi/abs/10.1086/259291 (Publisher: The University of Chicago Press) doi: 10.1086/259291

Bobba, M., Flabbi, L., & Levy, S. (2022). Labor Market Search, Informality, and Schooling Investments. *International Economic Review*, *63*(1), 211–259. Retrieved 2023-02-08, from https://onlinelibrary.wiley.com/doi/abs/10.1111/iere.12536 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/iere.12536) doi: 10.1111/iere.12536

Dix-Carneiro, R., Goldberg, P. K., Meghir, C., & Ulyssea, G. (2021, January). *Trade and Informality in the Presence of Labor Market Frictions and Regulations* [Working Paper]. National Bureau of Economic Research. Retrieved 2022-11-23, from https://www.nber.org/papers/w28391 doi: 10.3386/w28391

El Tiempo, C. E. (2023, December). El Éxito de Unicentro no es el único sellado por la Dian; van 355 cierres en el 2023. *El Tiempo*. Retrieved 2023-12-27, from https://www.eltiempo.com/economia/sectores/la-dian-cerro-355-establecimientos-en-el-2023-839025 (Section: economia)

Engbom, N. (2022, January). *Labor market fluidity and human capital accumulation* (Working Paper No. 29698). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w29698 doi: 10.3386/w29698

Heckman, J. J., Lochner, L., & Taber, C. (1998, January). Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents. *Review of Economic Dynamics*, *1*(1), 1–58. Retrieved 2024-01-05, from https://www.sciencedirect.com/science/article/pii/S1094202597900089 doi: 10.1006/redy.1997.0008

Jarosch, G., Oberfield, E., & Rossi-Hansberg, E. (2021). Learning From Coworkers. *Econometrica*, *89*(2), 647–676. Retrieved 2023-05-22, from https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA16915 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA16915) doi: 10.3982/ECTA16915

Lagakos, D., Moll, B., Porzio, T., Qian, N., & Schoellman, T. (2018, April). Life Cycle Wage Growth across Countries. *Journal of Political Economy*, *126*(2), 797–849. Retrieved 2023-04-27, from https://www.journals.uchicago.edu/doi/10.1086/696225 (Publisher: The University of Chicago Press) doi: 10.1086/696225

Lasry, J.-M., & Lions, P.-L. (2007, March). Mean field games. *Japanese Journal of Mathematics*, *2*(1), 229–260. Retrieved 2023-10-27, from http://link.springer.com/10.1007/s11537-007-0657-8 doi: 10.1007/s11537-007-0657-8

Ley 100. (1993). Retrieved from https://oig.cepal.org/sites/default/files/colombia_-_ley_100.pdf

Ley 3500. (1980). (Available at: https://www.bcn.cl/leychile/navegar?idNorma=7147&idParte=)

Lopez Garcia, I. (2015, April). *Human Capital and Labor Informality in Chile: A Life-Cycle Approach.* RAND Corporation. Retrieved 2023-04-03, from https://www.rand.org/pubs/working_papers/WR1087.html

Lopez-Martin, B. (2019, December). Informal Sector Misallocation. *Macroeconomic Dynamics*, *23*(8), 3065–3098. Retrieved 2022-11-04, from https://www.cambridge.org/core/journals/macroeconomic-dynamics/article/informal-sector-misallocation/9E3177CC56EB8EB49196F3D07CBC1900 (Publisher: Cambridge University Press) doi: 10.1017/S1365100517001055

Lucas, R. E., & Moll, B. (2014, February). Knowledge Growth and the Allocation of Time. *Journal of Political Economy*, *122*(1), 1–51. Retrieved 2023-02-20, from https://www.journals.uchicago.edu/doi/full/10.1086/674363 (Publisher: The University of Chicago Press) doi: 10.1086/674363

Lucas Jr, R. E. (2009). Ideas and Growth. *Economica*, *76*(301), 1–19. Retrieved 2023-08-08, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0335.2008

.00748.x (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0335.2008.00748.x) doi: 10.1111/j.1468-0335.2008.00748.x

Manuelli, R. E., & Seshadri, A. (2014, September). Human Capital and the Wealth of Nations. *American Economic Review*, *104*(9), 2736–2762. Retrieved 2023-11-02, from https://www.aeaweb.org/articles?id=10.1257/aer.104.9.2736 doi: 10.1257/aer.104.9.2736

Meghir, C., Narita, R., & Robin, J.-M. (2015, April). Wages and Informality in Developing Countries. *American Economic Review*, *105*(4), 1509–1546. Retrieved 2022-08-30, from http://www.aeaweb.org/articles?id=10.1257/aer.20121110 doi: 10.1257/aer.20121110

Perla, J., & Tonetti, C. (2014, February). Equilibrium Imitation and Growth. *Journal of Political Economy*, *122*(1), 52–76. Retrieved 2023-03-25, from https://www.journals.uchicago.edu/doi/10.1086/674362 doi: 10.1086/674362

Perla, J., Tonetti, C., & Waugh, M. E. (2021, January). Equilibrium Technology Diffusion, Trade, and Growth. *American Economic Review*, *111*(1), 73–128. Retrieved 2023-03-25, from https://pubs.aeaweb.org/doi/10.1257/aer.20151645 doi: 10.1257/aer.20151645

Perry, G. E., Arias, O., Fajnzylber, P., Maloney, W. F., Mason, A., & Saavedra-Chanduvi, J. (2007). *Informality: Exit and Exclusion.* The World Bank. Retrieved 2023-04-01, from http://elibrary.worldbank.org/doi/book/10.1596/978-0-8213-7092-6 doi: 10.1596/978-0-8213-7092-6

Samaniego De La Parra, B., & Fernández Bujanda, L. (2020, December). *Increasing the Cost of Informal Workers: Evidence from Mexico.* Banco de México. Retrieved 2023-10-23, from https://www.banxico.org.mx/publicaciones-y-prensa/documentos-de-investigacion-del-banco-de-mexico/%7B4F3B292C-1DB1-8A30-F63F-E789C9E0819A%7D.pdf doi: 10.36095/banxico/di.2020.19

Ulyssea, G. (2018, August). Firms, Informality, and Development: Theory and Evidence from Brazil. *American Economic Review*, *108*(8), 2015–2047. Retrieved 2023-02-15, from https://pubs.aeaweb.org/doi/10.1257/aer.20141745 doi: 10.1257/aer.20141745

Ulyssea, G. (2020a, January). *Formal and Informal Firm Dynamics.*

Ulyssea, G. (2020b, August). Informality: Causes and Consequences for Development. *Annual Review of Economics*, *12*(1), 525–546. Retrieved 2023-02-15, from https://www.annualreviews.org/doi/10.1146/annurev-economics-082119-121914 doi: 10.1146/annurev-economics-082119-121914

Appendix

A-1 Theory Appendix

A-1.1 Derivations

A-1.1.1 HJB equations

Consider first the case of a formal worker. For notational simplicity, let $g_f(z,t)$, the conditional probability density function for formal workers. That is, for every $z \in \Omega_f(t)$, then the probability density functions for formals is

$$g_f(z,t) = \frac{g(z,t)}{\mu_f(t)}$$

Analogously, denote by $g_i(z,t)$ the probability density function conditional on being an informal worker. Then, the value of being a formal worker after a time interval of length Δ is

$$V_f(z,t) = Y_f(z,t+\Delta)\Delta + T(t+\Delta)\Delta + \frac{1}{1+\rho\Delta}\left[\begin{array}{l} \alpha\Delta\mathbb{P}_f^f(t+\Delta)\int_{\Omega_f(t+\Delta)} \max\left\{V_f(\tilde{z},t+\Delta), V_f(z,t+\Delta)\right\} g_f(\tilde{z},t)\mathrm{d}\tilde{z} + \\ \alpha\Delta\mathbb{P}_f^i(t+\Delta)\int_{\Omega_i(t+\Delta)} \max\left\{V_i\left(\tilde{z},t+\Delta\right), V_f(z,t+\Delta)\right\} g_i(\tilde{z},t)\mathrm{d}\tilde{z} + \\ \left(1-(\alpha+\delta)\Delta\right)V_f(z,t+\Delta) \end{array}\right]$$

By multiplying both sides by $(1+\Delta\rho)$, dividing by Δ, and collecting terms, the Bellman equation can be rewritten as:

$$\begin{aligned} \frac{V_f(z,t) - V_f(z,t+\Delta)}{\Delta} &= Y_f(z,t+\Delta)(1+\rho\Delta) + T(t+\Delta) - (\rho+\delta)V_f(z,t+\Delta) \\ &+ \alpha\mathbb{P}_f^f(t+\Delta)\int_{\Omega_f(t+\Delta)} \max\left\{V_f(\tilde{z},t+\Delta) - V_f(z,t+\Delta), 0\right\} g_f(\tilde{z})\mathrm{d}\tilde{z} \\ &+ \alpha\mathbb{P}_f^i(t+\Delta)\int_{\Omega_i(t+\Delta)} \max\left\{V_i\left(\tilde{z},t+\Delta\right) - V_f(z,t+\Delta), 0\right\} g_i(\tilde{z})\mathrm{d}\tilde{z} \end{aligned}$$

By taking $\Delta \to 0$ yields the desired result. The derivation of the HBJ equation for informal workers is analogous.

A-1.1.2 Kolmogorov Forward Equation

Consider the CDF of the skill distribution G at time $t+\Delta$. For $z \leq \bar{z}$ the evolution can be described as

$$\begin{aligned} G(z,t+\Delta) &= \mathbb{P}\mathrm{r}\left(\tilde{z}_{t+\Delta} \leq z\right) \\ &= \int_0^z \underbrace{\left[1-\alpha\Delta\right]}_{\text{No Meeting}} g(\tilde{z},t) + \underbrace{\Delta\alpha\lambda_i^i(\bar{z}_{t+\Delta})G(z,t)}_{\text{Meeting below } z} g(\tilde{z},t)\mathrm{d}\tilde{z} \\ &= G(z,t) - \Delta\int_0^z \alpha g(\tilde{z},t)\mathrm{d}\tilde{z} + \Delta\int_0^z \alpha\lambda_i^i(\bar{z}_{t+\Delta})G(z,t)g(\tilde{z},t)\mathrm{d}\tilde{z} \end{aligned}$$

Thus by moving $G(z,t)$ to the left hand side and dividing by Δ it follows that

$$\frac{\partial G(z,t)}{\partial t} = \lim_{\Delta\to 0}\frac{G(z,t+\Delta)-G(z,t)}{\Delta} = -\alpha G(z,t)\left(1-G(z,t)\lambda_i^i(\bar{z}_t)\right) \tag{A-1}$$

Similarly, the same analysis follows for $z > \bar{z}_t$

$$\begin{aligned} G(z,t+\Delta) = &\int_0^{\bar{z}_t} \alpha\Delta\left[\lambda_i^f(\bar{z}_t)\left(G(z,t)-G\left(\bar{z}_t,t\right)\right)+\lambda_i^i(\bar{z}_t)G(\bar{z}_t,t)\right]g(\tilde{z},t)\mathrm{d}\tilde{z} \\ &+\int_{\bar{z}_t}^{z} \alpha\Delta\left[\lambda_f^f(\bar{z}_t)\left(G(z,t)-G\left(\bar{z}_t,t\right)\right)+\lambda_f^i(\bar{z}_t)G(\bar{z}_t,t)\right]g(\tilde{z},t)\mathrm{d}\tilde{z} \\ &\int_0^{z}\left[1-\alpha\Delta\right]g(\tilde{z},t)\mathrm{d}\tilde{z} \end{aligned}$$

By splitting the second term into two from 0 to $\bar{z}_t$ we can rewrite the past equation as

$$\begin{aligned} \frac{G(z,t+\Delta)-G(z,t)}{\Delta} = &\int_0^{\bar{z}_t} \alpha\left[\lambda_i^f(\bar{z}_t)\left(G(z,t)-G\left(\bar{z}_t,t\right)\right)+\lambda_i^i(\bar{z}_t)G(\bar{z}_t,t)-1\right]g(\tilde{z},t)\mathrm{d}\tilde{z} \\ &+\int_{\bar{z}_t}^{z} \alpha\left[\lambda_f^f(\bar{z}_t)\left(G(z,t)-G\left(\bar{z}_t,t\right)\right)+\lambda_f^i(\bar{z}_t)G(\bar{z}_t,t)-1\right]g(\tilde{z},t)\mathrm{d}\tilde{z} \end{aligned}$$

Then by using properties of our λ terms and taking the limit of Δ to 0 it yields the KFE for $z > \bar{z}_t$

$$\frac{\partial G(z,t)}{\partial t} = -\alpha\left(1-G(z,t)\right)\left[\lambda_i^f(\bar{z}_t)G(\bar{z}_t,t)+\lambda_f^f(\bar{z}_t)\left(G(z,t)-G(\bar{z}_t,t)\right)\right] \tag{A-2}$$

A-1.2 Proofs of propositions

A-1.2.1 Proof of proposition 1

Step 1: $V_i(0,t) > V_f(0,t)$ at any time t.

Step 2: $V_i(z,t)$ and $V_f(z,t)$ are monotonically increasing in z.

Step 3: $V_i(z,t)$ is concave if z.

Step 4: $V_f(z,t)$ is concave if z.

A-1.2.2 Proof of proposition 2

A-1.2.3 Proof of proposition 3

Note that using l'hospital rule on the Pareto assumption implies that

$$\lim_{x\to\infty}\frac{\Phi'(x)x}{x^{-\frac{1}{\theta}}} = \frac{k}{\theta}$$

Going back to the KFE for formal workers, dividing both sides of (**??**) by $x^{-1/\theta}$ yields

$$\gamma \frac{\Phi'(x)x}{x^{-\frac{1}{\theta}}} = \alpha \left(\frac{1 - \Phi(x)}{x^{-\frac{1}{\theta}}} \right) \left[\lambda_i^f(\bar{x})\Phi(\bar{x}) + \lambda_f^f(\bar{x}) \left(\Phi(x) - \Phi(\bar{x}) \right) \right]$$

Thus, by taking the limit as $x \to \infty$ and using our Pareto tail assumption (**??**), it follows that the growth rate of the economy is

$$\gamma = \alpha\theta \left[\lambda_i^f(\bar{x})\Phi(\bar{x}) + \lambda_f^f(\bar{x}) \left(1 - \Phi(\bar{x}) \right) \right] = \alpha\theta\psi \left(\Phi(\bar{x}) \right) \tag{A-3}$$

A-2 Empirical Appendix

Table A-1: Descriptive Statistics - ELCO

	2010	2013	2016	Overall	Obs
		Panel A: Cross-Sectional Sample			
Informal Status	0.592	0.658	0.633	0.631	19,120.000
	(0.491)	(0.474)	(0.482)	(0.483)	
Age	32.985	35.887	36.722	35.267	37,543.000
	(13.606)	(13.917)	(14.147)	(13.988)	
Wage (Real)	17,187.658	18,922.483	18,117.710	18,057.430	8,661.000
	(16,040.282)	(17,991.025)	(17,082.616)	(17,050.210)	
Female	0.563	0.546	0.544	0.550	37,543.000
	(0.496)	(0.498)	(0.498)	(0.497)	
Less High School	0.777	0.756	0.711	0.744	25,298.000
	(0.416)	(0.430)	(0.453)	(0.436)	
High School	0.050	0.058	0.070	0.061	25,298.000
	(0.218)	(0.234)	(0.255)	(0.239)	
2 Year College	0.086	0.113	0.134	0.115	25,298.000
	(0.281)	(0.316)	(0.341)	(0.319)	
4 Year College	0.087	0.073	0.085	0.080	25,298.000
	(0.281)	(0.261)	(0.279)	(0.272)	
		Panel B: Panel Sample			
Informal Status	0.597	0.643	0.615	0.621	16,206.000
	(0.491)	(0.479)	(0.487)	(0.485)	
Age	33.297	36.626	39.623	36.561	27,398.000
	(13.023)	(13.559)	(13.163)	(13.504)	
Wage (Real)	17,224.568	19,451.756	18,912.667	18,509.575	7,387.000
	(15,943.761)	(18,258.968)	(17,594.064)	(17,300.497)	
Female	0.581	0.563	0.567	0.570	27,398.000
	(0.493)	(0.496)	(0.496)	(0.495)	
Less High School	0.783	0.740	0.707	0.737	20,602.000
	(0.412)	(0.439)	(0.455)	(0.440)	
High School	0.048	0.061	0.070	0.062	20,602.000
	(0.213)	(0.240)	(0.256)	(0.241)	
2 Year College	0.089	0.119	0.135	0.118	20,602.000
	(0.285)	(0.324)	(0.341)	(0.323)	
4 Year College	0.080	0.080	0.088	0.083	20,602.000
	(0.272)	(0.271)	(0.284)	(0.276)	

Notes: Table A-1 shows the descriptive statistics of the key variables of the ELCA database (longitudinal survey of Colombia), for the years 2010, 2013 and 2016. Panel A is the sample of cross-sectional people who for who we have full information for the variables of informality status, age, education, and wages. Panel B is the sample of people followed in the three periods of analysis and with full information.

Table A-2: Descriptive Statistics - EPS

Variables	2004	2006	2009	2015	Overall	Obs
	Panel A - Cross-section sample					
Informality full	0.25	0.24	0.19	0.15	0.21	29,855
	(0.43)	(0.42)	(0.39)	(0.36)	(0.40)	
Age	38.76	34.62	37.70	38.44	37.34	29,855
	(11.44)	(11.57)	(11.45)	(12.11)	(11.66)	
Education level	2.18	2.33	2.32	2.31	2.29	29,855
	(0.97)	(1.00)	(0.98)	(0.78)	(0.93)	
Real Wage	5,001	5,364	5,576	6,875	5,740	29,855
	(4,541)	(4,790)	(4,687)	(5,203)	(4,818)	
Total experience (weekly hours)	24,067	23,083	24,394	23,785	23,822	29,855
	(15,044)	(14,320)	(14,196)	(15,481)	(14,759)	
Total informal experience	4,356	3,850	3,845	3,454	3,856	29,855
	(10,297)	(8,568)	(8,531)	(7,990)	(8,789)	
Total formal experience	19,711	19,233	20,549	20,331	19,966	29,855
	(14,714)	(14,229)	(14,184)	(15,192)	(14,582)	
Num jobs	4.02	4.38	4.46	3.95	4.20	29,855
	(2.57)	(2.67)	(2.70)	(2.70)	(2.66)	
Num informal jobs	0.83	1.00	0.94	0.83	0.90	29,855
	(1.47)	(1.56)	(1.50)	(1.40)	(1.49)	
Num formal jobs	3.16	3.34	3.48	3.09	3.27	29,855
	(2.44)	(2.52)	(2.51)	(2.45)	(2.48)	
	Panel B - Balance panel sample					
Informality panel	0.18	0.11	0.11	0.06	0.12	5,932
	(0.38)	(0.32)	(0.32)	(0.24)	(0.31)	
Age	35.26	37.28	39.76	46.47	39.69	5,932
	(8.93)	(8.93)	(8.95)	(8.94)	(8.94)	
Education level	2.21	2.25	2.26	2.20	2.23	5,932
	(0.92)	(0.93)	(0.94)	(0.78)	(0.89)	
Real Wage	4,557	5,240	5,514	7,253	5,641	5,932
	(3,590)	(4,075)	(4,301)	(5,074)	(4,260)	
Total experience (weekly hours)	41,124	41,124	41,124	41,124	41,124	5,932
	(10,125)	(10,125)	(10,125)	(10,125)	(10,125)	
Total informal experience	4,121	4,121	4,121	4,121	4,121	5,932
	(9,836)	(9,836)	(9,836)	(9,836)	(9,836)	
Total formal experience	37,003	37,003	37,003	37,003	37,003	5,932
	(11,989)	(11,989)	(11,989)	(11,989)	(11,989)	
Num jobs	6.09	6.09	6.09	6.09	6.09	5,932
	(2.64)	(2.64)	(2.64)	(2.64)	(2.64)	
Num informal jobs	0.87	0.87	0.87	0.87	0.87	5,932
	(1.67)	(1.67)	(1.67)	(1.67)	(1.67)	
Num formal jobs	5.20	5.20	5.20	5.20	5.20	5,932
	(2.45)	(2.45)	(2.45)	(2.45)	(2.45)	

Notes: Table A-2 shows the descriptive statistics of the key variables of the EPS database (Chile Social Protection Survey), for the years 2004, 2006, 2009, and 2015. Panel A is the sample of an unbalanced panel of people who have complete information for the variables of informality, age, education, and wages. Panel B is the sample of people followed in the four periods of analysis and with complete information.

A-3 Computational Appendix

A-3.1 Algorithm

We now describe the algorithm used to solve the equilibrium. Start with a guess $\bar{x}^{(0)}$. For the $n-$th step of the iteration

(1) Compute $\bar{\Phi}^{(n)}$.
Using equation (**??**) solve for $\bar{\Phi}$

$$1-\bar{\Phi}=\frac{[\lambda_i^f(\bar{\Phi})\bar{\Phi}+\lambda_f^f(\bar{\Phi})\left(1-\bar{\Phi}\right)]^2}{[\lambda_i^f(\bar{\Phi})\bar{\Phi}+\lambda_f^f(\bar{\Phi})\left(1-\bar{\Phi}\right)]\lambda_f^f(\bar{\Phi})+k(\bar{x}^{(n)})^{1/\theta}}$$

(2) Compute c_ℓ:
Using $\bar{x}^{(n)}$ and $\bar{\Phi}^{(n)}$ solve for c_ℓ using

$$\bar{\Phi}^{(n)}=\frac{c_\ell(\bar{x}^{(n)})^{\alpha/\gamma}}{1+\lambda_i^i(\bar{\Phi}^{(n)})c_\ell(\bar{x}^{(n)})^{\alpha/\gamma}}$$

Observe that the left-hand side is a constant between 0 and 1 while the right-hand side is a monotone-increasing function. Thus, if the solution for c_ℓ exists, it is unique.

(3) Solve Bellman equations: Consider the grid

$$\{x_1,...,x_N\}$$

and for notational convenience, denote $v_{ij}=v_i(x_j)$, $v_{fj}=v_f(x_j)$, and $\Phi(x_j)=\Phi_j$. The stationary Bellman equation for formal workers is

$$\begin{aligned}(\rho-\gamma)v_f(x)+\gamma x v_f'(x)&=(1-\tau)x-F\\&+\mathbb{1}_{\{x\geq\bar{x}\}}\left\{\alpha\lambda_f^f(\bar{x})\int_x^\infty\left(v_f(\tilde{x})-v_f(x)\right)\phi(\tilde{x})d\tilde{x}\right\}\\&+\mathbb{1}_{\{x<\bar{x}\}}\left\{\alpha\lambda_f^f(\bar{x})\int_{\bar{x}}^\infty\left(v_f(\tilde{x})-v_f(x)\right)\phi(\tilde{x})d\tilde{x}+\alpha\lambda_f^i(\bar{x})\int_x^{\bar{x}}\left(v_i(\tilde{x})-v_f(x)\right)\phi(\tilde{x})d\tilde{x}\right\}\end{aligned}$$

Similarly, the Bellman equation for informal workers is:

$$\begin{aligned}(\rho-\gamma)v_i(x)+\gamma x v_i'(x)&=x-\eta x^2\\&+\mathbb{1}_{\{x\geq\bar{x}\}}\left\{\alpha\lambda_i^f(\bar{x})\int_x^\infty\left(v_f(\tilde{x})-v_i(x)\right)\phi(\tilde{x})d\tilde{x}\right\}\\&+\mathbb{1}_{\{x<\bar{x}\}}\left\{\alpha\lambda_i^f(\bar{x})\int_{\bar{x}}^\infty\left(v_f(\tilde{x})-v_i(x)\right)\phi(\tilde{x})d\tilde{x}+\alpha\lambda_i^i(\bar{x})\int_x^{\bar{x}}\left(v_i(\tilde{x})-v_i(x)\right)\phi(\tilde{x})d\tilde{x}\right\}\end{aligned}$$

Define $h_j=x_j-x_{j-1}$ for all $j>1$, and we approximate the derivative of the value function as

$$v_f'(x_j)\approx\frac{v_j-v_{j-1}}{h_j}$$

Let $\bar{\ell}$ be the position in the grid such that if $j > \bar{\ell}$ then $x_j \geq \bar{x}$. Then, the discretized version of the Bellman equation for formal workers is

$$\left(\rho - \gamma + \frac{\gamma x_j}{h_j}\right) v_{fj} - \gamma x_j \frac{v_{f,j-1}}{h_j} = (1-\tau)x_j - F$$
$$+ \mathbb{1}_{\{j \geq \bar{\ell}\}} \left\{ \alpha\lambda_f^f \sum_{l=j}^{N} v_{fl}\phi_l h_l - \alpha\lambda_f^f \left(1 - \Phi_j\right) v_{fj} \right\}$$
$$+ \mathbb{1}_{\{j < \bar{\ell}\}} \left\{ \alpha\lambda_f^f \sum_{l=\bar{\ell}}^{N} v_{fl}\phi_l h_l - \alpha\lambda_f^f \left(1 - \bar{\Phi}\right) v_{fj} + \alpha\lambda_f^i \sum_{l=j}^{\bar{\ell}-1} v_{il}\phi_l h_l - \alpha\lambda_f^i(\bar{\Phi} - \Phi_j)v_j\right.$$

Similarly, for informal workers, the discretized version of the Bellman equation is

$$\left(\rho - \gamma + \frac{\gamma x_j}{h_j}\right) v_{ij} - \gamma x_j \frac{v_{i,j-1}}{h_j} = x_j - \eta x_j^2$$
$$+ \mathbb{1}_{\{j \geq \bar{\ell}\}} \left\{ \alpha\lambda_i^f \sum_{l=j}^{N} v_{fl}\phi_l h_l - \alpha\lambda_i^f \left(1 - \Phi_j\right) v_{ij} \right\}$$
$$+ \mathbb{1}_{\{j < \bar{\ell}\}} \left\{ \alpha\lambda_i^f \sum_{l=\bar{\ell}}^{N} v_{fl}\phi_l h_l - \alpha\lambda_i^f \left(1 - \bar{\Phi}\right) v_{ij} + \alpha\lambda_i^i \sum_{l=j}^{\bar{\ell}-1} v_{il}\phi_l h_l - \alpha\lambda_i^i(\bar{\Phi} - \Phi_j)v_{ij} \right\}$$

In addition, we impose the boundary condition $v_{f1} = v_{f0}$ and $v_{i1} = v_{i0}$. This determines a linear system on the vector $\vec{v} = [v_{i1}, \cdots, v_{iN}, v_{f1}, \cdots, v_{fN}]'$ described by the equation

$$\left[\begin{array}{c|c} A^i & C^f \\ \hline C^i & A^f \end{array}\right] \begin{bmatrix} v_{i1} \\ \vdots \\ v_{fN} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_{2N} \end{bmatrix}$$

where the matrices A^i, A^f, are $N \times N$ and its coefficients are defined by the formula

$$a_{lj}^i = \begin{cases} -\frac{x_j\gamma}{h_j} & \text{if } l = j-1 \\ \rho - \gamma + \dfrac{\gamma x_j}{h_j} + \alpha\lambda_i^f(1 - \Phi_j \mathbb{1}_{\{j > \bar{\ell}\}}) & \\ \qquad - \alpha\lambda_i^i(\Phi_j + \phi_j h_j)\mathbb{1}_{\{j<\bar{\ell}\}} + \alpha\bar{\Phi}[\lambda_i^i - \lambda_i^f]\mathbb{1}_{\{j<\bar{\ell}\}} & \text{if } l = j \\ -\alpha\lambda_i^i\phi_j h & \text{if } l < j < \bar{j} \end{cases},$$

$$a_{lj}^f = \begin{cases} -\frac{x_j\gamma}{h_j} & \text{if } l = j-1 \\ \rho - \gamma + \dfrac{\gamma x_j}{h_j} + \alpha\lambda_f^f \left((1 - (\Phi_j + \phi_j h_j)\mathbb{1}_{\{j \geq \bar{\ell}\}}\right) & \\ \qquad + \alpha\bar{\bar{\Phi}}[\lambda_f^i - \lambda_f^f]\mathbb{1}_{\{j<\bar{\ell}\}} - \alpha\lambda_f^i\Phi_j\mathbb{1}_{\{j<\bar{\ell}\}} & \text{if } l = j \\ -\alpha\lambda_f^f\phi_j h_j & \text{if } \bar{j} - 1 \leq l < j \end{cases},$$

the matrices C^i and C^f are also $N \times N$ and defined by

$$c_{lj}^i = \begin{cases} -\alpha\lambda_f^i \phi_j h_j & \text{if } l < j < \bar{j} \\ 0 & \text{otherwise} \end{cases} \qquad c_{lj}^f = \begin{cases} -\alpha\lambda_i^f \phi_j h_j & \text{if } \bar{j} - 1 \leq l < j \\ 0 & \text{otherwise} \end{cases}$$

and the return vector is

$$b_j = \begin{cases} x_j - \eta x_j^2 & \text{if } j \leq N \\ (1-\tau)x_{j-N} - F & \text{if } j > N \end{cases}$$

(4) Update $\bar{x}$ from indifference condition.