The value of a formal job

Adolfo Fuentes*
University of Edinburgh

Job Market Paper

September 15, 2024

Abstract

Starting a career with a formal or an informal job significantly affects labor market outcomes. This disparity arises from differences in productivity, human capital accumulation, job destruction rates, and the arrival rates of potential offers across sectors, alongside varying institutional settings such as minimum wage regulations. To explore these differences, I propose a novel model of a frictional labor market featuring two different sectors with varying institutional environments. Within each sector, firms differ in productivity and exogenous destruction rates, while workers engage in both off- and on-the-job search, accumulating human capital while employed. The above generates a job ladder where workers can improve their situation moving to a more productive, safer, or formal firm. I estimate the model using Chile's survey and administrative data from 2010-2019. The model shows that, on average, the value of a formal job is equivalent to a lump sum payment of 13.5 minimum wages, although is highly heterogeneous depending on the formal firm selected, and it is decreasing with the worker's level of human capital. Also, it shows that the consequences of starting with a formal job are present after 5 years, presenting differences in total earnings with a lower bound of 6.5%. Finally, I show that the primary components driving higher wages in the formal sector are greater productivity and better offer arrival rates, while enhanced job safety and faster human capital accumulation are secondary factors.

Keywords: Labor informality, job ladder, minimum wage, human capital accumulation.

*Email: a.i.fuentes@sms.ed.ac.uk. First version: August 2024. I am especially grateful to Axel Gottfries and Rafael Lopes de Melo for their continuous guidance and support throughout this project. I also thank Jesper Bagger, Mike Elsby, Jan Grobosek, Jonna Olson, the participants of the Job Market Session at the EALE Conference 2024, and the Macro Reading Group at the University of Edinburgh for their valuable comments. Finally, I am thankful for the financial support from the National Agency of Research and Development (ANID) of the Chilean Government for my doctoral studies. All errors are my own.

1 Introduction

Labor informality presents a pressing concern within developing economies, with far-reaching consequences for social welfare (Ulyssea, 2020). In these contexts, labor regulations do not extend to informal workers, exempting them from minimum wage requirements and regulated working hours. Furthermore, the informal sector generally evades taxation, which can strain government finances and also tend to sidestep contributions to the social security system, placing additional financial stress on health and pension programs (Levy & Schady, 2013). The scale of these challenges is considerable, with labor informality affecting a substantial portion of the employed population, ranging from 20% to 80% in developing countries (Perry et al., 2007; La Porta & Shleifer, 2014), and accounting for nearly one-third of total economic activity on average (Schneider & Enste, 2000; Elgin et al., 2021).

Past literature has shown that informal firms differ from formal ones in several aspects. First, formal firms tend to be more productive than informal ones (De Paula & Scheinkman, 2011). This can be explained as a result of a selection process, where more productive firms tend to hire more employees. As a consequence, they are more likely to be discovered and fined by the government, which in turn incentivizes them to operate formally (Meghir et al., 2015; Haanwinckel & Soares, 2021; Ruggieri & Cisneros-Acevedo, 2023; Parente, 2022). Second, informal firms tend to offer jobs with a higher unemployment risk (Perry et al., 2007; La Porta & Shleifer, 2008; De Paula & Scheinkman, 2011; La Porta & Shleifer, 2014). Third, previous research has demonstrated that human capital accumulation is greater in the formal sector compared to the informal one, which can be attributed to the availability of labor training and the acquisition of skills that are valuable to other employers (Lagakos et al., 2018; Bobba et al., 2021; Jedwab et al., 2023). Fourth, there is evidence showing that the transition rates across formal and informal jobs are heterogeneous depending on the characteristics of the job (Magnac, 1991; Meghir et al., 2015; Allen et al., 2018; Ulyssea, 2018). Consequently, workers in certain jobs are more or less likely to transition to the informal sector, and vice versa. Finally, the literature has shown that in the presence of search frictions in the labor market, informal firms create a negative externality for formal firms by making it more difficult for them to find workers, thereby increasing search costs (Ulyssea, 2018, 2020).

The literature has identified various ways in which informality can be problematic, and has studied each one of those from a separate perspective. However, there is limited research on the aggregate effects of these issues on workers' career paths, especially regarding the value of having a formal job and how this value evolves throughout a worker's career. Due to the substantial differences between sectors, how workers progress through their careers significantly impacts various outcomes, including wages, unemployment risk, human capital accumulation, and formality status. To help close this gap in the literature, I propose a frictional search model with two sectors

that differ in institutional settings. In this model, firms vary in productivity and exogenous job destruction rates inside each sector. Workers accumulate human capital while employed and lose it when unemployed. Additionally, workers can search for jobs both off- and on-the-job, allowing them to climb the job ladder by moving to more productive, safer, or formal firms.

The data used to calibrate the model will come from administrative records and employment survey data from Chile between 2010 and 2019. During this period, the informality rate in Chile remained relatively stable at around 28 percent of the labor force, formal employment accounted for approximately 65 percent, and the unemployment rate hovered around 7 percent (INE, 2019). The formal data consists of a 20 percent random sample from the administrative records of the unemployment insurance (*Seguro de Desempleo*). This sample includes workers registered in the most recent month of the database (currently June 2021) and features reconstructed work histories going backward. As is typical with such databases, public sector workers are not included in these records. The administrative records will be used to obtain information regarding the formal sector of the Chilean economy. The employment survey data corresponds to the National Employment Survey (*Encuesta Nacional de Empleo*) for the four quarters of the year, centered in February, May, August, and November. This data has a panel structure that follows individuals in a household over six quarters, and combining the entire data of those who participate in the labor market yields 1.9 million observations. Using survey data becomes necessary since it is the only way to have information relative to workers' characteristics in the informal sector, and the transitions between both sectors and unemployment. Finally, I also use the supplementary income survey (*Encuesta Suplementaria de Ingresos*) for the same period. This supplementary survey contains income data for those who answer the employment survey during the quarter centered in November each year, and it is the only source of income data for the informal sector. To calibrate the model I will match several moments from the data and the model.

With the calibrated model, it is possible to obtain the value of a formal and informal job in the firms operating in the labor market. It is important to notice that formal and informal jobs will take several values across the job ladder. Therefore, the extra value obtained from having a formal job concerning the informal counterpart will depend on the particular position of that job in the job ladder. This will generate a distribution of gains for formalization, allowing, for example, formal jobs to give more value than their informal counterparts at the beginning of the job ladder rather than at the top.

The paper is organized as follows: The second section reviews the contributions of this paper to the existing literature. The third section provides an overview of the Chilean context and details the data used. The fourth section introduces the model. The fifth section addresses the aspects related to model estimation. The sixth section presents the results of the calibration. The seventh

section discusses three exercises designed to study the value of holding a formal job. Finally, the eighth section concludes the paper and outlines potential directions for future research.

2 Related literature

This paper contributes to the literature studying labor informality from an structural perspective. Previous literature has focused on the effects of financial constraints in the labor decisions (Flabbi & Tejada, 2022), minimum wage effects over inequality and employment (Parente, 2022), different levels of human capital accumulation across sectors (Bobba et al., 2022), payroll effects on informality (Cisneros-Acevedo, 2022), complementarity between worker types (Haanwinckel & Soares, 2021), trade policy (Dix-Carneiro et al., 2021), the relative importance of the intensive and extensive margin Ulyssea (2018), and the effects of increasing regulation (Meghir et al., 2015). However, there is a gap in the literature regarding the impact of having a formal or informal job on a worker's career. The main objective of this paper is to help close this gap by incorporating the most relevant aspects of previous research into a job ladder model that includes an informal sector.

The contributions of this paper can be classified into three main components. First, it incorporates the between-employer competition and job ladder framework introduced by Cahuc et al. (2006), which was later extended to include human capital accumulation by Bagger et al. (2014), into the analysis of informal markets. This integration allows for a systematic understanding of the consequences of starting a career in either the formal or informal sector, taking into account differences in offer arrival rates, human capital accumulation, and exogenous destruction rates. In this context, workers with same levels of human capital and working in the same firm, but having a different benchmark option will receive different wages, increasing the range of wages observed in the economy. The resulting wider wage distribution and the possibility that more productive firms do not always pay higher wages than less productive ones can serve as a natural explanation for the overlap in wage distributions between the formal and informal sectors, as documented in the literature (Maloney, 1999; Hsieh & Olken, 2014; Allen et al., 2018; Ulyssea, 2018).

The second contribution relies on the inclusion of firms that can be different in three aspects: formality status, productivity and exogenous destruction rate.While the first two components are common in the literature (Haanwinckel & Soares, 2021; Bobba et al., 2022; Parente, 2022), differences in exogenous destruction rates within each sector have not been extensively studied in the literature on labor informality. This concept was recently introduced by Jarosch (2023) to examine the consequences of job loss, as it creates slippery job ladders at the lower end of the job distribution. Including this fact is relevant for several reasons: (i) Chilean formal firms experience annual destruction rates that vary by size, with differences of up to ten times (Arellano & Jimenez, 2016), (ii) It reduces the requirements to generate employment-to-employment transitions, since it

is not exclusively needed a higher productivity to move to another firm, (iii) Generates additional heterogeneity in the effective rates of human capital accumulation inside each sector, (iv) As I will show later, the empirical destruction rates decline with tenure, something that it is possible to explain with this setting.

The third contribution of this paper is the introduction of a minimum wage for the formal sector within a job ladder framework. Without a minimum wage, formal firms - being generally more productive, safer, and offering faster human capital accumulation - would likely pay lower wages than informal firms at the entry level of the job ladder. However, data shows that formal firms tend to offer higher wages across the entire wage distribution (INE, 2024), which can be attributed to the presence of a binding minimum wage. Additionally, between-employer competition partially extends the influence of the minimum wage to informal firms, as they must raise wages to reduce the risk of losing workers to similar formal sector firms through employment-to-employment transitions. Finally, the inclusion of a minimum wage in this setting allows to explore how the steady state equilibrium will react to changes in it.

3 Background and data

The period studied in this paper spans the years 2010 to 2019. This period is relatively stable in terms of labor market indicators, particularly with respect to unemployment and informality rates, which remained around 7 and 28%, respectively, during this time (Fuentes, 2019; INE, 2020).[1] Labor market participation at the national level was also relatively stable, averaging around 59%. However, we observe some internal shifts across gender. Men's participation slightly decreased from 72 to 70%, while women's participation significantly increased from 44 to 49% (INE, 2019). This change has been developing since the previous decade and can be primarily attribute to the expansion of tertiary education, cultural shifts, and the implementation of childcare policies during this period (Contreras & Plaza, 2007; Encina & Martínez, 2009; Contreras et al., 2012).

Regarding the minimum wage, it increased in roughly 50 percent in real terms between January 2010 and December 2019. Interestingly, this increase did not change either the ratio between the minimum and median wage in the economy, which was stable around 0.7 during 2010-2019, and the proportion of workers who earn the minimum wage in the formal sector, which fluctuated around 15 percent during the period according to the administrative records (Abud et al., 2022), suggesting that the minimum wage policy during this period was innocuous in this aspect.

[1]Official data on informality has only been available since 2017; however, in another paper (Fuentes, 2024) I compare formal sector employment from survey data with administrative records, and show that the informality rate in the country remained constant throughout the decade of 2010.

The definition of a formal job is closely tied to the institutional context of the labor market in question, making it essential to clarify what a formal job entails in Chile. The Chilean National Bureau of Statistics (INE) provides an institutional definition that includes criteria from both the worker's and firm's perspectives. For a job to be considered formal, it must involve a written contract, offer vacation time, pay at least the minimum wage, and contribute to social security on the worker's side. On the firm's side, the company must issue a document, typically a pay slip, detailing the worker's gross wage along with deductions for social security contributions, taxes, and insurance. Additionally, the firm must be registered with the tax office (INE, 2017).

Unfortunately, data corresponding to this definition has only been available since 2017. Therefore, I use an alternative definition of labor informality based on whether the firm contributes to social security, particularly the pension system. The pension system requires a contribution of 10 percent of the gross wage, which is deposited into an individual account, along with an additional commission of approximately 1 percent that goes to the pension fund administrator. This information has been available since 2010, enabling the analysis of informality over the entire decade. This definition captures a combined condition involving both the worker and the firm: to receive social security contributions, the worker must have a labor contract, and the firm must be registered in the social security system and, by extension, with the tax office. While this definition is less stringent than the one used by the INE, it is widely employed in the literature on informality (Ulyssea, 2020).

The data used in this paper comes from two distinct datasets. First, I utilize administrative records from the unemployment insurance system, which covers 20 percent of randomly selected workers in the formal sector, excluding the public sector. The sample is drawn by randomly selecting 20 percent of workers from the most recent month of the dataset (June 2021) and reconstructing their labor histories back to 2002, the year when the unemployment insurance system began. This dataset includes IDs for both workers and firms, and provides information on labor income at the individual and firm levels, as well as on gender, education, citizenship, age, region, and economic sector.

An important consideration is that individuals contribute to the unemployment insurance system for each formal job they hold. Consequently, individuals with multiple formal jobs will appear multiple times in the records each month. To address this, I will focus on the primary employment for each individual in the analysis, where primary employment is defined as the job with the highest wage among other criteria. The specific procedure for determining the primary employment is detailed in the Appendix.[2] As the dataset tracks individuals over time, it allows for the identification of employment spells in the formal sector. After structuring the data as a panel, following the procedure outlined in the Appendix, the dataset will contain approximately 243 million obser-

[2]The procedure can be seen in Appendix 1: Main job in administrative data

vations from 2.2 million individuals.[3]

A key assumption in using the administrative data is that workers not employed in the formal sector are classified as unemployed rather than as employed in the informal sector. This assumption facilitates the calculation of human capital loss during unemployment periods. As will be demonstrated later, this assumption is reasonable given the low transition rates from the formal to the informal sector.

To analyze the informal sector, I will use panel data from the national employment survey. This survey tracks households over six quarters and collects information on employment status, age, gender, education, job tenure, region, economic sector, and occupation, though it does not include income data. Each worker is assigned a unique ID, allowing for tracking across the six quarters. While I cannot fully reconstruct individual labor histories, I can follow their employment history from their first observed job. Also, it is important to mention that since I see workers between quarters, I cannot distinguish between employment-to-employment transitions and workers who lost their jobs and then find another one between quarters. Then, I will assume that workers who exhibit two different jobs in two consecutive quarters did an employment-to-employment transition. The survey is conducted monthly, with each dataset covering data from the past three months. I will focus on data from February, May, August, and November. There are 40 surveys from February 2010 to November 2019, with each survey containing around 47,000 worker observations, resulting in approximately 1.88 million total observations.

To complement the employment data, I will use the supplementary income survey, which is conducted during the last quarter of the year (around November). This survey provides detailed information on various measures of labor income and uses an ID that matches the one in the employment survey, allowing for the linkage of income data with employment transitions. This linkage is crucial, as it is the only source of informal income data connected to employment information. To ensure consistency with the approach used for administrative records, I will focus solely on the primary employment. This method will result in approximately 423,000 observations.

As is common in labor economics literature, I will restrict my sample to full-time workers who are unlikely to change their level of education. Specifically, I will focus on individuals aged 25 to 64 who are part of the labor force. Additionally, I will include only those workers with a self-reported job tenure of 17 years or less and who report working between 30 and 60 hours per week. For unemployed individuals, I will limit the sample to those who have been seeking a job for less than three years. Several filters were also applied to ensure consistency in the reported tenure and

[3]The procedure can be seen in Appendix 2: Administrative data to panel

sector, with the details provided in the Appendix.[4]

Table 1 presents summary statistics for the sample used in the analysis, including the overall sample as well as breakdowns by gender and education level (tertiary education versus secondary education or less). Regarding employment and unemployment shares, nearly 68 percent of workers are employed in the formal sector, 21 percent in the informal sector, and the remaining 11 percent are unemployed. The shift in proportions compared to the full dataset reflects the exclusion of partially employed individuals, which increases the unemployment share. In terms of gender differences, men are slightly more likely to be employed in both sectors, while women show a higher propensity for unemployment. Interestingly, workers with tertiary education also participate in the informal sector, though their participation is, as expected, lower than that of those with secondary education or less.

Table 1: Summary statistics for employment: 2010-2019

	Overall	Men	Women	Tertiary	Non-tertiary
Share (%) - Formal Employ.	67.6	68.2	66.6	72.2	65.1
Share (%) - Informal Employ.	21.5	22.1	20.7	16.8	24.0
Share (%) - Unemployment	10.9	9.6	12.7	11.1	10.9
Tenure (Months) - Formal Job	31.6	30.1	33.7	35.2	29.5
Tenure (Months) - Informal Job	28.5	28.5	28.4	31.1	27.5
Search (Months) - Unemployment	4.6	4.2	5.1	5.6	4.1
Share (%) - Men	57.1	-	-	51.1	60.2
Share (%) - Women	42.9	-	-	48.9	39.8
Share (%) - Tertiary education	34.1	30.5	38.8	-	-
Share (%) - Secondary or less	65.9	69.5	61.2	-	-
Observations	819,110	468,899	350,211	278,987	540,123

Source: Data from ENE surveys between 2010 and 2019.

In terms of tenure, the formal sector shows higher job tenure compared to the informal sector. When comparing genders, men tend to have lower tenure in the formal sector but experience fewer months of unemployment than women. When the sample is divided by education, workers with tertiary education degrees have longer tenures in both sectors, but they also experience longer periods of unemployment, a pattern documented in Fuentes (2019).

Table 2: Summary statistics for income: 2010-2019

	Overall	Men	Women	Tertiary	Non-tertiary
Real log wage (Monthly) - Formal Job	13.1	13.1	13.0	13.4	12.9
Real log wage (Monthly) - Informal Job	12.6	12.7	12.5	13.0	12.5
Observations	161,953	94,424	67,529	52,369	109,584

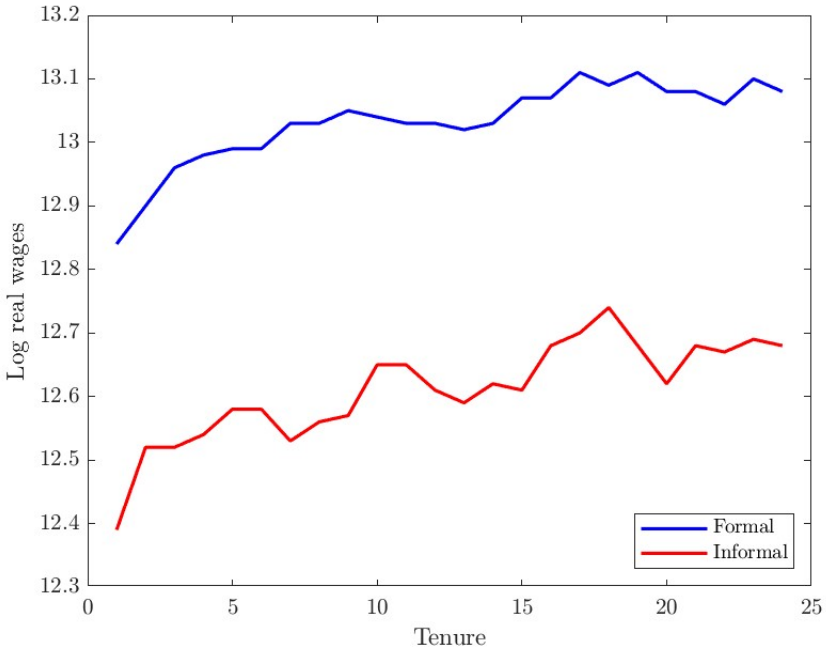
Source: Data from ESI merged with ENE surveys between 2010 and 2019.

[4]This procedure can be seen in Appendix 3: Filters to ensure consistency in tenure

For the wage analysis, I use the employment dataset after applying the previously mentioned filters and merge it with wage information. I then calculate real wages using November 2019 as the baseline. Additionally, I focus on workers between the 6th and 95th percentiles of the wage distribution for each sector. Table 2 summarizes the income data. As shown, wages in the formal sector are higher than in the informal sector. Additionally, men earn higher wages than women in both sectors. In addition, there appears to be a wage premium associated with higher education in both sectors. It is also worth noting that these statistics for the formal sector align with those found in the administrative dataset. Summary statistics for tenure and wages in the administrative data can be found in the Appendix.[5]

To conclude this section, I will analyze how wages and unemployment risk evolve with job tenure in the dataset. Specifically, I calculate the average real log wage and employment-to-unemployment transition rates for each level of tenure during the first two years of employment. As shown in Figure 1, wages in the formal sector start higher than those in the informal sector, and wages in both sectors tend to increase with tenure. However, an interesting observation is that even after two years, wages in the informal sector do not reach the levels seen in the formal sector during the first months of employment.

Figure 1: Evolution of real log wage per hour by tenure and sector (2010-2019)



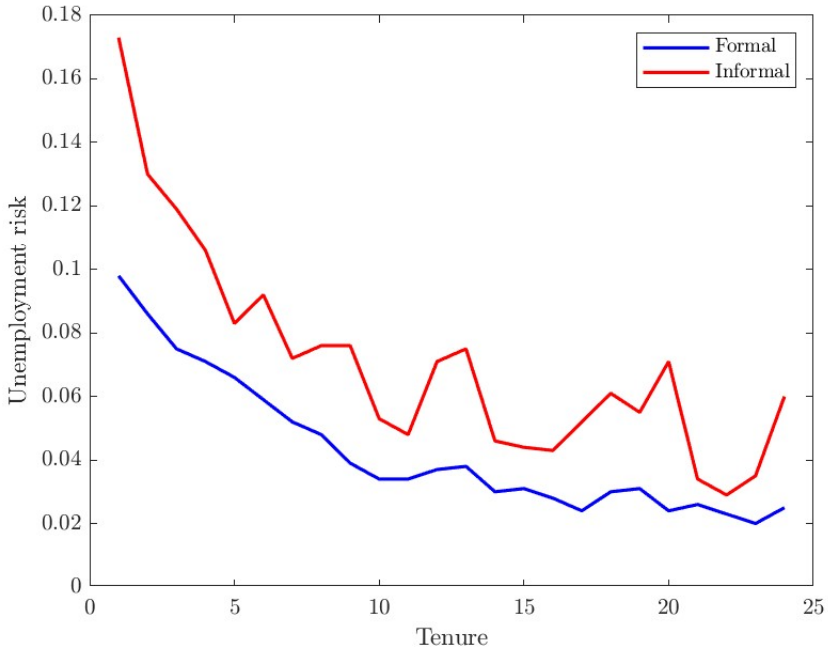
Source: Data from ESI surveys between 2010 and 2019.

Figure 2 illustrates the evolution of employment destruction rates across tenure in both sectors. As shown, informal jobs are consistently riskier than formal jobs, and both sectors experience a

[5]Table in Appendix 4: Descriptive statistics administrative data

decline in unemployment risk with increasing tenure. An interesting observation from this figure is that a worker with two years of tenure in the informal sector may face a lower unemployment risk than a newly hired worker in the formal sector. This is important because it suggests that workers at the start of the job ladder may opt to move to the informal sector if the safety gains outweigh the productivity differences.

Figure 2: Employment to unemployment transitions by tenure and sector (2010-2019)



Source: Data from ENE surveys between 2010 and 2019.

4 Model

4.1 Setting

Time is assumed to be discrete. Firms are characterized by their compliance status (formal or informal) and a vector $\theta = (\theta_y, \theta_\delta)$, where θ_y represents the firm's productivity and θ_δ denotes the exogenous job destruction rate associated with the firm. Workers are infinitely lived, and all of them search randomly for offers, samplying them from the same distribution. Workers can be in three possible states: unemployed (u), employed in the informal sector (i), and employed in the formal sector (f). Formal firms differ from the informal firms in two aspects: i) They sample their productivity and destruction rate parameters from different distributions, and ii) Formal firms must pay at least the minimum wage ($\overline{w}$).

When workers are unemployed, they receive a value z and they encounter job openings in formal and informal sector with probabilities λ_f and λ_i, respectively. When they are formally employed, they encounter job openings in formal and informal sectors with probabilities λ_{ff} and λ_{fi}. And, when they are informally employed, they encounter job openings in formal and informal sectors with probabilities λ_{if} and λ_{ii}.

Workers are endowed with a level of human capital h, which is observable to the firms and where $h \in \mathcal{H} = [\underline{h}, \overline{h}]$. Human capital is assumed to evolve following a process $G_e(h'|h)$ where $e \in \{f, i, u\}$. The above implies that human capital accumulation will differ in each possible state.

Once a worker and a firm form a match, output is determined by $p(\theta_y, h)$. If the match is dissolved, either because the worker transitions to another firm or due to exogenous destruction, the job opening holds no continuation value for the firm.

The timing of the model is as follows: First, the output $p(\theta_y, h)$ is observed and wages are paid. Second, workers may become unemployed due to the firm's exogenous destruction rate θ_δ. Third, based on their new state, human capital is updated according to the process $G_e(h'|h)$. Fourth, if workers remain employed, they may receive outside offers with sector-specific probabilities $(\lambda_{ff}, \lambda_{fi}, \lambda_{if}, \lambda_{ii})$ and can use these offers to renegotiate with their current firm or move to a new one. If workers are unemployed, they can receive offers from each sector with probabilities λ_f and λ_i.

In this model wages are restricted to fixed contracts, but can be renegotiated when workers receive an outside offer or experience an increase in their human capital. The wage determination for the informal sector follows the framework of Cahuc et al. (2006) as adapted in Jarosch (2023). In the formal sector, a modified version of this bargaining process is proposed to account for the minimum wage ($\overline{w}$).

4.2 Bargaining process

The bargaining process in this model has three main components. A first component includes the characteristics of the current firm in which a worker is employed, including compliance, productivity, and destruction rate denoted by c and θ. A second relevant component involves the characteristics of the best alternative offer received up to that point, or benchmark, which will be denoted by $\hat{c}$ and $\hat{\theta}$. The existence of a benchmark is relevant to generate the between-employer competition as presented in Cahuc et al. (2006) and Bagger et al. (2014). It is also important to note that the benchmark encompasses all possible firms in each sector, as well as unemployment. In this context, unemployment is treated as a "match" that provides an income of z and cannot

be destroyed. Finally, the third component is the level of human capital h.

Let $W(c, \hat{c}, \theta, \hat{\theta}, h)$ denote the value of an employed worker in a firm with compliance status c, vector θ, and human capital h, and where $\hat{c}$ and $\hat{\theta}$ represent the sector and characteristics of the benchmark firm. Then $J(c, \hat{c}, \theta, \hat{\theta}, h)$ represents the value of a filled job, and $U(h)$ denote the value of unemployment.

With these components, it is possible to define the surplus of a match, $S(c, \theta, h)$, as:

$$S(c, \theta, h) \equiv max\{W(c, \hat{c}, \theta, \hat{\theta}, h) - U(h) + J(c, \hat{c}, \theta, \hat{\theta}, h), 0\}$$

It can be shown that this surplus depends solely on the current levels of θ and h, in addition to the compliance status c. The intuition behind this result is similar to that presented in Jarosch (2023): the benchmark firm's parameters determine the distribution of the surplus, not the total amount of surplus generated. However, in this model with two sectors, the surplus also depends on the compliance status of the current firm. This is because the compliance status affects the probability of receiving future offers in each sector and because the minimum wage ($\overline{w}$) is applicable only in the formal sector.

If an unemployed worker with human capital h matches with an informal firm characterized by θ, the value of the employment match will be such that the worker receives a share α of the surplus, where α represents the worker's bargaining power. Then:

$$W(i, u, \theta, u, h) - U(h) = \alpha S(i, \theta, h)$$

This structure implies that workers will prefer firms with higher surpluses, as they will receive a greater amount when transitioning from unemployment to employment.

If an unemployed worker with human capital h matches with a formal firm, the same procedure applies as described above. However, if the worker's share α of the surplus is insufficient to meet the minimum wage, it is assumed that the worker's bargaining power increases to $\gamma(\theta, h)$. Here, $\gamma(\theta, h)$ represents the portion of the surplus that must be allocated to a worker with human capital h by a formal firm with characteristics θ to ensure that the minimum wage $\overline{w}$ is being paid. Then:

$$W(f, u, \theta, u, h) - U(h) = \max\{\alpha, \gamma(\theta, h)\} S(\theta, h)$$

Where the match will only occur if $0 < \gamma(\theta, h) < 1$.

This structure assumes that workers in both the formal and informal sectors will have the same

bargaining power when the minimum wage is not binding. Although this assumption may not be evident, it is commonly used in the literature on labor informality to ensure that results are not skewed by differences in bargaining power across workers (Haanwinckel & Soares, 2021).

Next, consider a worker with human capital h currently employed at an informal firm with characteristics (i, θ), and using a benchmark firm with characteristics $(\hat{c}, \hat{\theta})$. If the worker remains with the firm (i, θ) rather than the benchmark firm $(\hat{c}, \hat{\theta})$, it implies that $S(i, \theta) > S(\hat{c}, \hat{\theta})$, otherwise, the worker would have chosen the benchmark firm. In this context, the between-employer competition implies that the current firm will outcompete the benchmark firm. To achieve this, the current firm will provide the worker with the entire surplus offered by the benchmark firm $S(\hat{c}, \hat{\theta})$, plus a share α of the surplus difference between the two firms $S(i, \theta, h) - S(\hat{c}, \hat{\theta}, h)$. Then:

$$W(i, \hat{c}, \theta, \hat{\theta}, h) - U(h) = S(\hat{c}, \hat{\theta}, h) + \alpha[S(i, \theta, h) - S(\hat{c}, \hat{\theta}, h)]$$

For a worker with human capital h currently employed at a formal firm with characteristics (f, θ), and using a benchmark firm with characteristics $(\hat{c}, \hat{\theta})$, the bargaining follows a similar logic, but we must account for the minimum wage $\overline{w}$. We need to consider two scenarios. In the first case, where the minimum wage is not binding, the worker receives:

$$NB \quad = \quad S(\hat{c}, \hat{\theta}, h) + \alpha[S(f, \theta, h) - S(\hat{c}, \hat{\theta}, h)]$$

In the second case, the minimum wage is binding, and therefore, the worker receives:

$$B \quad = \quad \gamma(\theta, h) S(f, \theta, h)$$

Then considering both cases, the value of being employed at a formal firm is the maximum of the two:

$$W(f, \hat{c}, \theta, \hat{\theta}, h) - U(h) = max\{NB, B\}$$

Regarding the decision process, consider a worker employed at a firm with characteristics (c, θ) and a benchmark $(\hat{c}, \hat{\theta})$ who receives an offer from a firm with characteristics (c', θ'). The procedure is as follows: If (c', θ') is a formal firm, we need to compute $\gamma(\theta', h)$. If $0 < \gamma(\theta', h) < 1$ the offer is considered feasible, otherwise the match is not possible. For a feasible offer, whether from the formal or informal sector, we have three cases to examine:

1. If $S(c', \theta', h) > S(c, \theta, h)$, the worker moves to θ'. The set of all options that result in a movement is defined as M. Then, in the case (c', θ', h) is informal the new value of

employment is given by:

$$W(i, c, \theta', \theta, h) - U(h) = S(c, \theta, h) + \alpha \left[S(i, \theta', h) - S(c, \theta, h) \right]$$

And in the case (c', θ', h) is formal:

$$W(f, c, \theta', \theta, h) - U(h) = max\{NB, B\}$$

Where each condition is given by:

$$\begin{aligned} NB &= S(c, \theta, h) + \alpha[S(f, \theta', h) - S(c, \theta, h)] \\ B &= \gamma(\theta', h) S(f, \theta', h) \end{aligned}$$

2. If $S(c, \theta, h) > S(c', \theta', h) > S(\hat{c}, \hat{\theta}, h)$ then the worker re-bargains, changing the benchmark from $(\hat{c}, \hat{\theta})$ to (c', θ'). The set of all options that result in a re-bargain is defined as RB. Then, the value of employment will be given by:

$$\begin{aligned} W(i, c, \theta, \theta', h) - U(h) &= S(c, \theta', h) + \alpha \left[S(i, \theta, h) - S(c, \theta', h) \right] \\ W(f, c, \theta, \theta', h) - U(h) &= max\{NB, B\} \end{aligned}$$

Where each condition is given by:

$$\begin{aligned} NB &= S(c, \theta', h) + \alpha[S(f, \theta, h) - S(c, \theta', h)] \\ B &= \gamma(\theta, h) S(f, \theta, h) \end{aligned}$$

Note that this process means that whenever a worker re-bargains using a better benchmark in the informal sector, it results in an improved value of employment. However, this is not always the case for a worker in the formal sector. For example, if the minimum wage is binding for the worker and they receive an offer from another firm where the minimum wage is also binding, but this new firm offers a better benchmark, the worker would retain the current value of employment. In this situation, only the benchmark changes, with no impact on the value of employment.

3. If $S(c, \theta, h) > S(\hat{c}, \hat{\theta}, h) > S(c', \theta', h)$ then nothing happens, and the worker keeps the current firm and benchmark.

4.3 Value functions

Given the setting mentioned above, then the value of unemployment is given by:

$$U(h) = z + \beta \int_{\mathcal{H}} \Bigg(\lambda_f \int_M [W(f,u,x,u,h') - U(h')]\, dF_f(x) \ ... \\ + \lambda_i \int_M [W(i,u,x,u,h') - U(h')]\, dF_i(x) + U(h') \Bigg) dG_u(h'|h)$$

The intuition behind this equation is as follows: At the beginning of a period, the unemployed worker receives the value z. In the next period, there is a probability λ_f of receiving a formal job offer. Any offer that provides a value higher than the current value of unemployment (equivalent to a positive surplus) will be accepted. When a worker transitions from unemployment to employment, the benchmark is unemployment. Similarly, with probability λ_i, the worker may receive an informal offer, which follows the same process. Additionally, there is a possibility of not receiving any offer and remaining unemployed into the next period, where the worker faces a risk of losing human capital, as determined by the process $G_u(h'|h)$.

The value of being employed in the formal sector is given by:

$$W(f,\hat{c},\theta,\hat{\theta},h) = \max\{w(f,\hat{c},\theta,\hat{\theta},h), \overline{w}\} + \beta \int_{\mathcal{H}} [(1-\theta_\delta) \times ... \\ \Bigg(\lambda_{ff} \Bigg(\int_M W(f,f,x,\theta,h')dF_f(x) + \int_{RB} W(f,f,\theta,x,h')dF_f(x) \Bigg) + ... \\ \lambda_{fi} \Bigg(\int_M W(i,f,x,\theta,h')dF_i(x) + \int_{RB} W(f,i,\theta,x,h')dF_i(x) \Bigg) \Bigg) + ... \\ \Bigg(1 - \lambda_{ff} \int_{M \cup RB} dF_i(x) - \lambda_{fi} \int_{M \cup RB} dF_f(x) \Bigg) \times W(f,\hat{c},\theta,\hat{\theta},h) ... \\ +\theta_\delta U(h')]\, dG_f(h'|h)$$

A worker employed in the formal sector will receive the higher of either the wage function or the minimum wage ($\overline{w}$). In the next period, if the job is not exogenously destroyed, the worker may receive an offer from either the formal or informal sector. Given such an offer, the worker can choose to move to a new job or use the offer to renegotiate their current condition. If no offer is received, the worker remains with the same firm in the following period. If the worker continues employed, there is a chance of increasing their human capital according to the updating process of the formal sector. However, if the job is exogenously destroyed, the worker begins the next period in unemployment.

Analogously, the value of employment in the informal sector is given by:

$$
\begin{aligned}
&W(i,\hat{c},\theta,\hat{\theta},h) = w(i,\hat{c},\theta,\hat{\theta},h) + \beta \int_{\mathcal{H}} [(1-\theta_\delta) \times ... \\
&\left(\lambda_{if}\left(\int_M W(f,i,x,\theta,h')dF_f(x) + \int_{RB} W(i,f,\theta,x,h')dF_f(x)\right) + ... \right. \\
&\left.\lambda_{ii}\left(\int_M W(i,i,x,\theta,h')dF_i(x) + \int_{RB} W(i,i,\theta,x,h')dF_i(x)\right)\right) + ... \\
&\left(1 - \lambda_{fi}\int_{M\cup RB} dF_i(x) - \lambda_{ii}\int_{M\cup RB} dF_f(x)\right) \times W(i,\hat{c},\theta,\hat{\theta},h)... \\
&+\theta_\delta U(h')]\, dG_i(h'|h)
\end{aligned}
$$

The primary difference between this value function and the one for the formal sector is that the probabilities of receiving offers shift from the pair $(\lambda_{ff}, \lambda_{fi})$ to $(\lambda_{if}, \lambda_{ii})$. Additionally, there is no minimum wage constraint in the informal sector.

Regarding firms, we have that the value of a formal filled vacancy is given by:

$$
\begin{aligned}
&J(f,\hat{c},\theta,\hat{\theta},h) = p(\theta_y,s) - \max\{w(f,\hat{c},\theta,\hat{\theta},h),\overline{w}\} + ... \\
&\beta\int_{\mathcal{H}}(1-\theta_\delta)\left(\lambda_{ff}\int_{RB} J(f,f,\theta,x,h')dF_f(x) + ... \right. \\
&\lambda_{fi}\int_{RB} J(f,i,\theta,x,h')dF_i(x) \;+\; ... \\
&\left.\left(1 - \lambda_{ff}\int_{M+RB} dF_f(x) - \lambda_{if}\int_{M+RB} dF_i(x)\right) \times\; J(f,\hat{c},\theta,\hat{\theta},h')\right) dG_f(h'|h)
\end{aligned}
$$

In this scenario, a formal firm receives the value of production and pays the higher of either the unconstrained wage or the minimum wage. If the job is not exogenously destroyed in the next period, the worker may receive offers. If the worker chooses to move to a new job, the vacancy is terminated and holds no continuation value for the firm. The firm retains value from the filled vacancy only when the worker either uses the new offers to renegotiate or receives no offers at all. Additionally, if the worker remains employed, there is a probability that their level of human capital will increase.

Following the same reasoning, the value of a filled vacancy in the informal sector is given by:

$$
\begin{aligned}
&J(i,\hat{c},\theta,\hat{\theta},h) = p(\theta_y,s) - w(i,\hat{c},\theta,\hat{\theta},h) + ... \\
&\beta \int_S (1-\theta_\delta) \Big(\lambda_{if} \int_{RB} J(i,f,\theta,x,h') dF_f(x) + ... \\
&\lambda_{ii} \int_{RB} J(i,i,\theta,x,h') dF_i(x) + ... \\
&\left(1 - \lambda_{if} \int_{M+RB} dF_f(x) - \lambda_{ii} \int_{M+RB} dF_i(x) \right) \times J(i,\hat{c},\theta,\hat{\theta},h') \Big) \, dG_i(h'|h)
\end{aligned}
$$

As with the value of employment, the key differences are the absence of a minimum wage and the probabilities of receiving offers from the informal sector.

Next, the last group of value functions needed in order to solve the model are the surplus functions. This functions represent the total gains obtain from the match are then are going to be divided between the firm and the worker, according to the bargaining rules. Following the definition of surplus, it can be shown that the surplus in the formal sector is given by:

$$
\begin{aligned}
&S(f,\theta,h) = p(\theta_y,s) - z + \beta \int_H (1-\theta_\delta) \left(S(f,\theta,h') + ... \right. \\
&\lambda_{ff} \int_M \left[\max\{ S(f,\theta,h') + \alpha[S(f,x,h') - S(f,\theta,h)], \gamma(x,h') S(f,x,h') \} ... \right. \\
&\left. - S(f,\theta,h') \right] dF_f(x) + \lambda_{fi} \int_M \left[\alpha[S(i,x,h') - S(f,\theta,h')] \right] dF_i(x) \Big) \, dG_f(h'|h)) ... \\
&- \beta \int_H \Big\{ \lambda_f \int_M [\max\{ \gamma(\theta,h') S(f,x,h') dF_f(x) \} + ... \\
&\lambda_i \int_M \alpha S(i,x,h') dF_i(x)] \Big\} \, dG_u(h'|h) + ... \\
&\beta \int_H U(h') dG_f(h'|h) - \beta \int_H U(h') dG_u(h'|h)
\end{aligned}
$$

As mentioned earlier, the surplus does not depend on the benchmark values $(\hat{c}, \hat{\theta})$.

Conceptually, first it captures the difference between the production at the current job and unemployment flows. Second, it shows the difference in gains from doing on-the-job versus off-the-job search. This difference will depend significantly on which state provides more access to offers and the quality of those offers. This process involves comparing the pair (λ_f, λ_i) with $(\lambda_{ff}, \lambda_{fi})$ and $(\lambda_{if}, \lambda_{ii})$. The intuition is that if the pair (λ_f, λ_i) strictly dominates the other two, workers will demand better jobs since accepting a job reduces their chances of receiving superior offers later on. Conversely, if (λ_f, λ_i) is strictly dominated by the other two, workers may accept lower-quality

offers, recognizing that they will gain more search opportunities while employed.

Third, the surplus also accounts for human capital accumulation gains in the formal sector compared to unemployment. Since the literature indicates that human capital tends to erode during unemployment, this difference is likely to be positive.

Finally, the surplus only depends on the movement set (M) and not on the re-bargain set (RB). This result, again, is intuitive since re-bargaining does not increases the surplus, just the way in which is divided between parts.

Then, the surplus for the informal sector is given by:

$$
\begin{aligned}
&S(i,\theta,h) = p(\theta_y, s) - z + \beta \int_H (1-\theta_\delta)\,(S(i,\theta,h') + ... \\
&\lambda_{ii} \int_M \left[\alpha[S(\tilde{i},x,h') - S(f,\theta,h')]\right] dF_i(x) + ... \\
&\lambda_{if} \int_M \Big[\max\{S(i,\theta,h') + \alpha[S(\tilde{f},x,h') - S(i,\theta,h)], \gamma(x,h')S(\tilde{f},x,h')\}... \\
&-S(f,\theta,h')\Big]\, dF_f(x))\, dG_i(h'|h) + ... \\
&-\beta \int_H \Big\{\lambda_f \int_M [\max\{\gamma(\theta,h')S(f,x,h')dF_f(x)\} + ... \\
&\lambda_i \int_M \alpha S(i,x,h')dF_i(x)]\Big\}\, dG_u(h'|h) + ... \\
&\beta \int_H U(h')dG_i(h'|h) - \beta \int_H U(h')dG_u(h'|h)
\end{aligned}
$$

4.4 Surplus properties and wages

As stated in Jarosch (2023), these functional equations can be solved jointly with the bargaining protocol and the value of unemployment. A key property of the surpluses is that they summarize the value of a job by considering the three-dimensional ladder components: sector, productivity, and destruction rate. Given a set of model parameters, value function iteration can be used to determine the surpluses that generate a steady-state equilibrium in the model, allowing for the simulation of labor histories. However, there is a difference between the procedure outlined in Jarosch (2023) and the one presented here. In the former, wage computations are unnecessary for obtaining surpluses, whereas in this model, wage calculations are needed to ensure that formal firms are complying with the minimum wage requirement.

5 Estimation

5.1 Functional assumptions and strategy

To solve the model, I will make some additional assumptions to ensure tractability. First, I assume there are only five productivity types and five destruction rate types in each sector. This creates 25 firm types per sector, resulting in 50 total firm types. The productivity and destruction rate for each firm will be randomly drawn from independent Beta distributions, with separate distributions for the formal and informal sectors. These four Beta distributions will be discretized. Regarding the output function, the functional form will be additively separable, $p_c(\theta, h) = \underline{p}_c + s + \theta_y$, where there will be two shifters, one for each sector.

Human capital will be discretized into five levels, i.e, $h \in \{1, 5\}$. The updating of human capital will follow a Markov process, where workers have a probability ψ_f of increasing their human capital in one unit in the formal sector and a probability ψ_i in the informal sector. Additionally, unemployed workers will experience a decrease in their human capital in one unit with probability ψ_u.

With these assumptions, the set of parameters to estimate is given by Table 3.

Table 3: Parameters to estimate

Parameter	Description
λ_f	Formal offer arrival during unemployment
λ_i	Informal offer arrival during unemployment
λ_{ff}	Formal offer arrival during formal employment
λ_{fi}	Informal offer arrival during formal employment
λ_{if}	Formal offer arrival during informal employment
λ_{ii}	Informal offer arrival during informal employment
$\eta_{\delta f}, \mu_{\delta f}$	Unemployment risk distribution formal sector
$\eta_{\delta i}, \mu_{\delta i}$	Unemployment risk distribution informal sector
α	Bargaining power of workers
η_{yf}, μ_{yf}	Productivity distribution formal sector
η_{yi}, μ_{yi}	Productivity distribution informal sector
$\underline{p}_f$	Common output shifter formal sector
$\underline{p}_i$	Common output shifter informal sector
ψ_f	Skill accumulation during formal employment
ψ_i	Skill accumulation during informal employment
ψ_u	Skill depreciation during formal unemployment
z	Flow value of unemployment

For the estimation strategy, I will follow Jarosch (2023) and apply the Simulated Method of Moments with indirect inference to estimate the model parameters on a monthly basis, including moments to match for the informal sector. The method minimizes the difference between the

observed (or indirectly inferred) moments in the data and those generated by the model simulations. The model is estimated in a steady state, meaning that the flows across different states are balanced.

5.2 Moments

The model has 21 parameters and is calibrated using 26 moments. The first two moments (1-2) capture unemployment-to-employment (UE) transitions into the formal and informal sectors. These moments correctly identify the arrival rates of offers, denoted by λ_f and λ_i.

The next four moments (3-6) are associated with the employment-to-employment (EE) transitions between sectors. These moments are monotonically related to the parameters λ_{ff}, λ_{fi}, λ_{if}, and λ_{ii}.

The next two moments (7-8) are related with the average job destruction rate in each sector. These moments inform the first shape parameter of the Beta distributions in each sector $\eta_{\delta f}$ and $\eta_{\delta i}$. To inform about the second shape parameter, I compute six moments (9-14) on how the unemployment risk changes through tenure in the labor market. To obtain this, I run the following regression for each sector:

$$I_{it}^{EU} = \alpha_0 + \sum_{\tau=1}^{\tau_{max}} \beta_\tau D_{it}^{\tau} + X_{it} + \varepsilon_{it}$$

Where I_{it}^{EU} is a binary variable indicating whether worker i transitioned to unemployment in period t. The variable D_{it}^{τ} is a binary variable indicating whether the worker has been employed at the same firm for D consecutive quarters. The coefficients β_τ reflect how job tenure reduces the destruction rate faced by workers. The matrix X includes a set of control variables such as gender dummies, month and year fixed effects, age group dummies (16-24, 25-34, 35-44, 45-54, and 55-64), education categories, and interactions between gender and those categories.

Then I will compute the average of the β_τ coefficients for the first, second, fourth and seventh year of tenure. Then, the moments used to calibrate the parameters $\mu_{\delta f}$ and $\mu_{\delta i}$ will be the differences between the averages of the last three groups and the first one:

$$\frac{1}{4}\sum_{\tau=5}^{8}\hat{\beta}_\tau - \frac{1}{4}\sum_{\tau=1}^{4}\hat{\beta}_\tau, \quad \frac{1}{4}\sum_{\tau=13}^{16}\hat{\beta}_\tau - \frac{1}{4}\sum_{\tau=1}^{4}\hat{\beta}_\tau, \quad \frac{1}{4}\sum_{\tau=25}^{28}\hat{\beta}_\tau - \frac{1}{4}\sum_{\tau=1}^{4}\hat{\beta}_\tau$$

Regarding wages, I will jointly calibrate the parameters α, $\underline{p}_f$, $\underline{p}_i$, and the productivity pairs in both sectors (η_{yf}, μ_{yf}) and (η_{yi}, μ_{yi}) using ten moments (15-24). The first pair are the mean log

wages in each sector. The second and third pairs represent the differences between log wages at the 90th and 50th percentiles, and the 50th and 10th percentiles, in both sectors. The fourth pair captures the annual wage growth in each sector. The final pair reflects the wage gap in both sectors, defined as:

$$\mathrm{G}_c = 1 - \frac{\overline{w}_c^0}{\overline{w}_c}$$

Where $\overline{w}_c^0$ is the average wage for those starting a job after unemployment in the sector c, and $\overline{w}_c$ is the average wage in that sector.

Although all moments are influenced by the aforementioned parameters, some parameters affect certain moments more than others, which provides intuition about their relationships. The average log wages primarily inform the output shifters and the first shape parameter of both productivity distributions. The differences in log wages across percentiles help identify the second shape parameters of the productivity distributions. Meanwhile, the wage growth rates and wage gaps in both sectors are informative about bargaining power and both parameters of the productivity distributions.

For the loss of human capital while unemployed, I will estimate the following regression:

$$w_{it}^0 = \alpha_0 + \gamma_1 \tau_{it}^u + \zeta_1 \overline{w}_i + \varepsilon_{it}$$

Where w_{it}^0 is the log wage after an unemployment spell, τ_{it} are the months of duration of the unemployment spell, and $\overline{w}$ is the average log wage of the worker. Then the moment $\hat{\gamma}_1$ (25) relates monotonically to the parameter ψ_u. It is important to note that, due to the nature of the data for the informal sector, it is not possible to accurately estimate this regression using data from that sector. However, since the model assumes that unemployment affects workers from both the formal and informal sectors equally, this parameter can be estimated using only administrative data from the formal sector.

For the accumulation of human capital, I will run the following regression:

$$w_{it}^t = \alpha_0 + \gamma_2 \tau_{it}^t + \zeta_2 w_{it} + \varepsilon_{it}$$

Where w_{it}^t is the log wage after t periods of continued tenure, τ^t is measuring those months, and w_{it} is the wage at the first period of employment in that firm. Then the moments γ_2 (26-27) identify the components ψ_f and ψ_i. Here, it is possible to estimate the parameters in both sectors with the data available.

Finally, for the parameter z, I will use the observed minimum wage in the informal sector as moment (28) to calibrate it. The intuition is that when workers decide to take a job, they are forgoing the flow z, making this parameter relevant to determine the indifference point between working and not working. Since the model includes a minimum wage in the formal sector, that sector cannot be used to assess the opportunity cost of working. However, this is not the case of the informal sector, and therefore the minimum wage observed there is informative of z.

5.3 Firms in equilibrium

The algorithm used to solve the model can be found in the Appendix.[6] Although the model has a fixed number of firms, not all of them necessarily operate in equilibrium. This outcome can be driven by two factors. First, a formal firm may not participate in the market if, given the production parameters ($\underline{p}_f$, η_f, and μ_f), it is unable to pay at least the minimum wage. In the case of an informal firm, it may be absent from the market equilibrium if, given its production parameters ($\underline{p}_i$, η_i, and μ_i), it cannot offer a wage high enough to compensate unemployed workers for forgoing the flow z.

5.4 Simulation

Once the model has assigned a set of parameters, it will simulate labor histories for 10,000 workers over 420 months (representing 35 years). The workers will start unemployed with the lowest level of human capital and will receive offers and make employment decisions accordingly. The moments will be computed using the entire simulated dataset, aligning with the procedure applied to the real data. It is important to note that although the event of receiving an offer is random across simulations for the same worker, the firm associated with that offer is fixed across simulations. This means that if a worker, say W, receives an offer in the first period, it will come from the same firm in both simulations, even though W might not receive an offer at all in some simulations. This approach ensures that differences in the estimated parameters are not attributable to random variations in the simulations.[7]

6 Calibration

The moments used and the calibrated parameters can be seen in Table 4. First of all, it is important to mention that the overall calibration has a good fit. In general terms, the fit is very good for the informal sector, while it has some differences with the formal one.

[6] Details on the algorithm Appendix 6: Algorithm

[7] More details about the simulation can be found in Appendix 7: Simulation details.

Table 4: Parameters and moments used to calibrate model

Moment	Target	Model	Parameter
Job finding rate - Formal	9.5%	9.5%	$\lambda_f = 0.095$
Job finding rate - Informal	5.7%	5.7%	$\lambda_i = 0.068$
E-E Transitions - FF	1.5%	0.7%	$\lambda_{ff} = 0.037$
E-E Transitions - FI	0.4%	0.2%	$\lambda_{fi} = 0.021$
E-E Transitions - IF	2.1%	0.4%	$\lambda_{if} = 0.028$
E-E Transitions - II	2.3%	0.4%	$\lambda_{ii} = 0.094$
Mean job loss - Formal	1.3%	2.0%	$\eta_{\delta,f} = 0.906$
Mean job loss - Informal	2.9%	3.0%	$\eta_{\delta,f} = 0.937$
Decay destruction - Formal (2 years)	-1.3%	-0.3%	$\mu_{\delta,f} = 0.718$
Decay destruction - Formal (4 years)	-1.6%	-0.6%	
Decay destruction - Formal (7 years)	-1.8%	-0.8%	
Decay destruction - Informal (2 years)	-2.2%	-0.5%	$\mu_{\delta,i} = 0.375$
Decay destruction - Informal (4 years)	-2.6%	-0.8%	
Decay destruction - Informal (7 years)	-2.7%	-0.9%	
Average log wages - Formal	13.02	13.18	$\underline{p}_f = 1.250$
Average log wages - Informal	12.63	12.84	$\underline{p}_i = 0.687$
$w_{90} - w_{50}$ - Formal	0.965	0.647	$\eta_{y,f} = 0.656$
$w_{50} - w_{10}$ - Formal	0.503	0.394	$\mu_{y,f} = 0.187$
$w_{90} - w_{50}$ - Informal	0.845	0.501	$\eta_{y,i} = 1.125$
$w_{50} - w_{10}$ - Informal	0.857	0.482	$\mu_{y,f} = 1.250$
Wage growth - Formal	7.9%	6.5%	
Wage growth - Informal	18.1%	16.9%	
Wage gap - Formal	20.2%	23.9%	$\alpha = 0.450$
Wage gap - Informal	24.1%	34.6%	
Human capital acc - Formal	0.19%	0.19%	$\psi_f = 0.131$
Human capital acc - Informal	0.16%	0.16%	$\psi_i = 0.100$
Human capital destruction	-0.30%	-0.30%	$\psi_u = 0.184$
Minimum wage in formal sector	11.10	10.98	$z = 12.11$

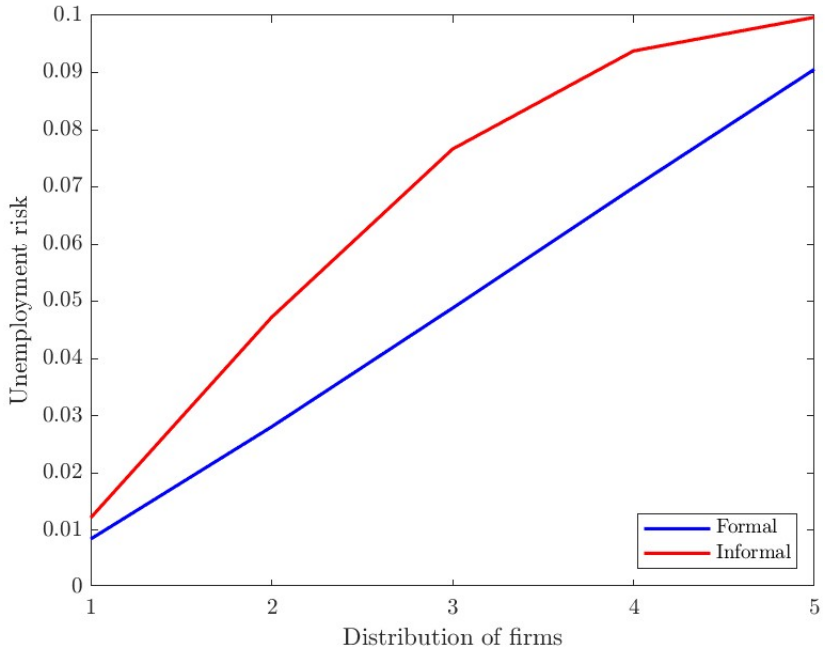
Regarding the job finding rates, the model successfully replicates the moments observed in the data. The fact that $\lambda_f > \lambda_i$, indicating that unemployed workers find it easier to secure a formal job than an informal one, and that $\lambda_{if} < \lambda_f$, meaning it is easier to find a formal job while unemployed than when employed in the informal sector, implies a higher value of unemployment relative to being employed in the informal sector. Because of the above, workers entering the informal sector will require significant compensation to offset the reduced likelihood of transitioning to the formal sector. However, it is important to note that while the model preserves the ranking of offer arrival rates observed in the data, it only partially explains the employment-to-employment transitions, particularly within both sectors.

About the moments associated with the destruction rate distributions, the model overestimates the average job loss for the formal sector, although it correctly captures that the mean destruction

rate is significantly higher in the informal sector compared to the formal sector. As for the decline in unemployment risk, the model demonstrates that the risk decreases more rapidly in the informal sector, though the rate of decline is below the one observed in the data.

An interesting aspect of the unemployment risk distribution is the significant overlap between the two sectors, as illustrated in Figure 3, which depicts the discretized Beta distributions for unemployment risk in both sectors. This suggests that the differences between formal and informal firms are not primarily driven by employment stability. This finding aligns with Arellano & Jimenez (2016), who show considerable dispersion in the annual destruction rates of firms within the formal sector, ranging from 1.3 to 12.0 percent in 2013. In the Chilean context, the main differences between sectors instead stem from productivity distributions and output shift parameters.

Figure 3: Distribution of destruction rates across types of firms

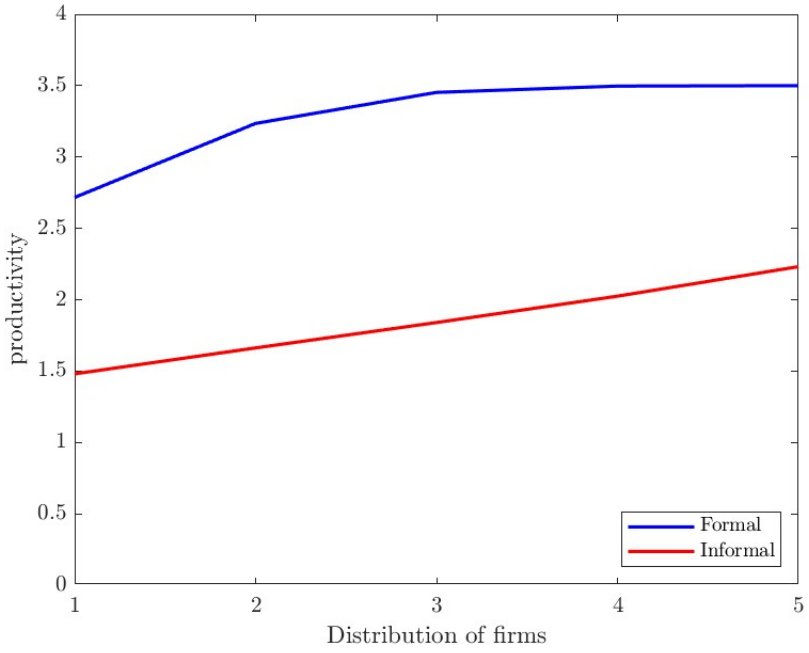


For the wage-related moments, the model shows a relatively good fit for average wages. In terms of wage percentile differences, the model performs qualitatively well, capturing the higher wage dispersion above the median in the formal sector and the more balanced dispersion in the informal sector. However, these differences are smaller than those observed in the data. Regarding annual wage growth in both sectors, the model replicates the qualitative trend that wage growth is higher in the informal sector, though there are discrepancies in the magnitudes. For wage gaps, the model correctly captures the higher wage gap in the informal sector, but the differences between the model and the data are significant, particularly in the informal sector.

Regarding the parameters associated with wages, the output shifter in the formal sector ($\underline{p}_f$)

is considerably higher than in the informal sector ($\underline{p}_i$). In contrast, formal sector productivity is more concentrated, while productivity in the informal sector is more dispersed, as shown in Figure 4. The estimated bargaining power is relatively low compared to what is typically found in the literature (for example, Jarosch (2023) found $\alpha = 0.92$ for Germany). Unfortunately, there are no comparable studies on bargaining power in Chile to further contextualize this result.

Figure 4: Distribution of productivity across types of firms



Regarding the human capital upgrading process, the model accurately matches the moments observed in the data. As expected, $\psi_f > \psi_i$, indicating that human capital accumulation is faster in the formal sector, which aligns with the literature (Bobba et al., 2021). This suggests that the primary advantage of formal sector employment is access to more productive firms and higher rates of human capital accumulation. The human capital destruction rate, ψ_u, is greater than both accumulation parameters, implying that destruction occurs more rapidly than accumulation.

About the minimum wage observed in the informal sector, the model accurately assesses that this value is significantly below the minimum wage in the formal sector (12.61). The fact that the value of z (12.11) is higher than the minimum wage observed in the informal sector in the model (10.98) implies that workers are willing to accept wages below their reservation flow because they anticipate accumulating human capital and receiving better offers in the future.

7 The value of a formal job

With the calibrated model, it is possible to estimate the value of a formal job. Since the question of what constitutes the value of a formal job can be interpreted in various ways, this section will explore different approaches to addressing it.

First, I will take a theoretical approach by comparing the value functions for workers employed in both sectors. Second, I will run a simulation where half of the workers are guaranteed an offer from a formal firm, while the other half receive an offer from an informal firm, and I will compare labor outcomes 1, 5, and 10 years later. Lastly, I will present an exercise to understand the relative importance of different aspects of the formal and informal sector in the total earnings.

7.1 Differences in value functions

Using the value functions for employed workers provides a direct way to compare the value of formal and informal jobs. However, a few important considerations must be kept in mind. The results are presented in units of the minimum wage in 2019.[8] Second, these value functions apply to infinitely lived individuals, meaning they reflect the present discounted value of an infinite stream of wages. Third, the selection of firms to compare is relevant. For this exercise, I will compare the value of being employed at a formal and an informal firm that occupy the same relative position in their respective distributions. This approach is preferable to forcing an informal firm to operate formally, as a significant proportion of these firms would exit the market if required to pay the minimum wage.

However, even in the scenario described, it is not clear which value functions to compare, as the value of being employed at a firm depends on the benchmark used in bargaining. To simplify the exercise, I will compare the value of being employed between firms when the benchmark is unemployment. This approach is chosen for two reasons. First, the value of being employed with a benchmark is not always valid, as it requires the benchmark to have a lower surplus than the current firm. This means that firms with higher surpluses will offer fewer comparable value functions. Second, since on-the-job offer arrivals are relatively infrequent, most workers in both sectors will initially use unemployment as their benchmark during the early periods of employment.

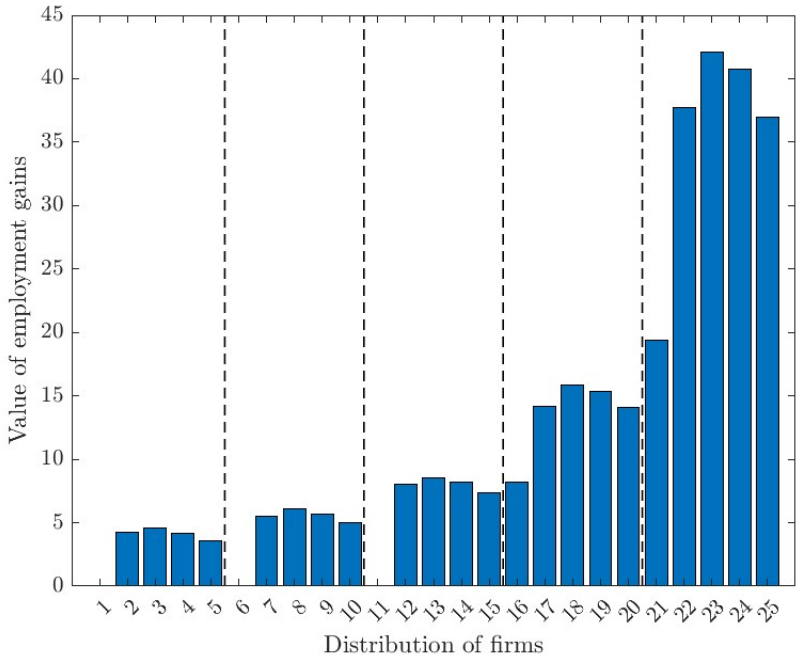
Since the value of being employed depends on human capital levels, I calculate the differences in value for each level. The results for the first level of human capital are shown in Figure 5. This figure is structured to display $W(f, \theta, u, u, 1) - W(i, \theta, u, u, 1)$ for all θ. There are 25 possible pairs $(\theta_y, \theta_\delta)$. The figure arranges firms from those with higher unemployment risk to the safer ones.

[8] The value was CLP \$301,000 at that time, which is equivalent to approximately USD \$435 today.

Within each risk group, the first firm is the least productive, and the last is the most productive. If the corresponding informal firm is not in the market, I assign a value of 0 for the difference. This applies to firms 1, 6, and 11, which are the least productive with the three highest levels of unemployment risk.

First, the figure demonstrates that transitioning from an informal to a formal firm increases the value of employment. This increase is primarily driven by access to more productive firms, the minimum wage, lower destruction rates, faster human capital accumulation, and better offer arrival rates in the formal sector. On average, the gains from moving to a formal firm are equivalent to a lump sum payment of 14.3 minimum wages. However, this value increases significantly when considering only the top four firms, averaging 39.4 minimum wages.

Figure 5: Gains of moving to a formal job for workers with $h = 1$



As expected, safer firms generate higher gains. For example, firms 2, 7, 12, 17, and 22 all share the same productivity component, but their unemployment risk decreases progressively. Interestingly, gains do not always increase with productivity. Consider firms 21, 22, 23, 24, and 25. While they share the same destruction rate and have increasing productivity, the surplus is highest in firm 22. This occurs because the value of being employed takes into account search gains, i.e., the probability of receiving offers from better firms, which is higher in firms with lower surpluses. However, these firms also tend to be riskier or less productive, creating a trade-off between search gains and the firm's conditions. In this specific case, the trade-off is maximized in the middle of the distribution.

Second, the gains from holding a formal job decrease as human capital levels increase. This occurs because workers at higher levels on the human capital ladder derive less incremental value from the faster accumulation of human capital. To illustrate this, I will compute the average gains of moving to a formal job across the five levels of human capital.[9] Table 5 shows the results. The gains from holding a formal job are on average 13.5 minimum wages, approximately USD$ 5,870. The gains decrease gradually up to the third level of human capital. For the fourth level, the gains are 7% lower than at the first level, and they increase to 20% at the fifth level.

Table 5: Average gains of having a formal job by level of human capital

Level of human capital (h)	Average gains of a formal job	Relative to $h = 1$
1	14.34	1.000
2	14.28	0.996
3	14.05	0.980
4	13.34	0.930
5	11.51	0.802
Average	13.50	0.941

7.2 Simulations

Regarding the simulations, the procedure will be as follows: I will simulate an economy with 10,000 individuals, starting as unemployed with the lowest level of human capital. Half of these individuals will receive an offer from a formal firm, while the other half will receive an informal offer. Offers will come from firms that are in equilibrium and will be at the same relative position in the joint distribution of productivity and destruction rate across sectors. I will consider two cases to represent the lowest and median of this joint distribution. Since these firms have positive surpluses, workers will accept the offers and start working at these firms. I will then compute the total earnings, average wage, and unemployment periods after 1, 3, and 5 years. The results are can be seen in Table 6.

In the table, earnings and average wages are measured in units of the minimum wage. For the lowest productivity firms, workers receive around 23% higher earnings and average wages in the formal sector during the first year. However, the difference in unemployment periods is practically negligible, as the destruction rates of both firms are similar. Therefore, the gains in higher earnings are primarily due to being employed at a more productive firm.

For the median firms, there are noticeable differences in both average wages (17%) and unemployment periods (-25%), attributed to variations in productivity and destruction rates between

[9]Here, I will only consider firms that are in the market for those with the first level of human capital to avoid generating additional gains for non-existent firms.

the firms. Additionally, the average wage in the median formal firm is lower than that in the lowest productivity formal firm. This occurs because firms with lower destruction rates, internalize that they provide higher search and human capital gains, and therefore they can pay lower wages.

Table 6: Labor outcomes from simulation after 1,3 and 5 years.

	1 Year - Lowest			1 Year - Median		
	Formal	Informal	Difference	Formal	Informal	Difference
Earnings	13.46	10.85	24.0%	14.11	11.40	23.8%
Average wage	1.40	1.13	23.3%	1.34	1.15	16.9%
Unemployment periods	2.39	2.40	-0.4%	1.54	2.04	-24.7%
	3 Years - Lowest			3 Years - Median		
	Formal	Informal	Difference	Formal	Informal	Difference
Earnings	42.96	38.87	10.5%	46.16	39.75	16.1%
Average wage	1.49	1.34	11.0%	1.51	1.35	12.3%
Unemployment periods	7.33	7.08	3.6%	5.87	6.62	-11.3%
	5 Years - Lowest			5 Years - Median		
	Formal	Informal	Difference	Formal	Informal	Difference
Earnings	76.41	71.74	6.5%	79.85	72.58	10.0%
Average wage	1.54	1.45	6.6%	1.57	1.45	8.0%
Unemployment periods	10.78	10.62	1.5%	9.45	10.15	-6.9%

After 3 years, the differences between the two scenarios are smaller, though still significant. Earnings are 10.5% higher for those who started at the lowest productivity firms, and up to 16.1% higher for those who started at the median productivity firms. The decrease in differences is expected, as workers eventually leave their initial firms, either due to unemployment shocks, which lead them to sample new offers, or because they move to different firms.

In the last row, we observe the results of the simulation 5 years after the initial job allocation. It is noteworthy that, despite the time elapsed, the differences in earnings remain significant: 6.5% for those who started at the lowest productivity firms and 10% for those who started at the median productivity firms.

In addition to the exercise presented here, I repeat the simulation assigning workers to the firms with the highest productivity and lowest destruction rates in both sectors. As expected the differences in labor outcomes are highly significant, showing that after 5 years, the differences in total earnings are close to 42%. This exercise can be seen in the Appendix.

Overall, these results suggests that starting a career in a formal job has substantial long-term value, as the differences in outcomes persist over the medium term, with a minimum difference

around 6.5% in total earnings after 5 years.

7.3 Relative importance

The final exercise involves recreating the simulations for the median firm while altering various parameters to understand their impact on total earnings after 5 years. The main differences between the formal and informal sectors can be attributed to four factors: different productivity distributions, represented by the pairs (η_{yf}, μ_{yf}) and (η_{yi}, μ_{yi}); varying rates of job offers both off- and on-the-job, determined by λ_f, λ_i, λ_{ff}, λ_{fi}, λ_{if}, and λ_{ii}; different destruction rates, represented by the pairs $(\eta_{\delta f}, \mu_{\delta f})$ and $(\eta_{\delta i}, \mu_{\delta i})$; and finally, different rates of human capital accumulation, given by ψ_f and ψ_i.

For each exercise, I will compute a new equilibrium of the model by imposing that one group of parameters associated with the formal sector be equal to their counterparts in the informal sector. After obtaining this new equilibrium, I will repeat the simulation for workers starting in the median firm and compute their total earnings after 5 years.

The results of this exercise are presented in Table 7. As seen, total earnings decrease for those who receive an offer in the formal sector. This occurs because the informal sector's parameters are generally less favorable than those in the formal sector. Interestingly, workers who receive an informal offer also experience a decline in earnings. This result is not only driven by poorer outcomes when transitioning to the formal sector but also due to a drop in wages within the informal sector. The intuition behind this is that since the informal sector offers worse conditions, such as limited search gains due to lower offer arrival rates and slower human capital accumulation, informal firms must offer compensating differentials to attract workers. After the parameter changes, the formal sector becomes more like the informal one, reducing the compensating differentials and leading to lower wages in the informal sector.

After performing the exercises, it is evident that the primary differentiating parameters between the sectors are those related to the productivity distributions. When these parameters are equalized, earnings in the formal sector decrease by nearly 21%, while in the informal sector, they decline by 28%. The second most significant factor is the arrival of offers; when this parameter is equalized across sectors, earnings drop by 13% in the formal sector and 18.5% in the informal sector.

The parameters with the least relative importance are unemployment risk, which causes a 6.3% change in earnings in the formal sector, and human capital accumulation, which results in a change of just under 3%.

Table 7: Total earnings after changes in structural parameters

Scenario	Formal		Informal	
Baseline	79.85		72.58	
Change	Earnings	Diff.	Earnings	Diff.
Same human capital acc.	77.55	-2.9%	70.22	-3.3%
Same unem. risk	74.85	-6.3%	68.85	-5.1%
Same productivity	63.19	-20.9%	52.09	-28.2%
Same transitions	69.41	-13.1%	59.16	-18.5%

8 Conclusion and further research

The literature on labor informality has explored various factors that distinguish the formal sector from the informal one. Key characteristics differentiating formal and informal firms include varying levels of productivity, unemployment risk, and faster human capital accumulation in the formal sector. However, there remains a gap in understanding how these components, when combined, influence workers' career trajectories. This paper addresses this gap by examining the value of a formal job.

To answer this question, I develop a novel model with search frictions that incorporates a three-dimensional job ladder (sector, productivity, and unemployment risk). In this model, workers accumulate human capital at different rates depending on the sector and engage in both off- and on-the-job search. Additionally, formal sector firms are required to pay their workers at least the minimum wage.

I calibrate the model using both administrative and survey data from Chile between 2010 and 2019. During this period, Chile exhibited stable employment trends across the formal and informal sectors. Overall, the model achieves a reasonable fit, though certain components, particularly those related to the arrival of on-the-job offers and job destruction rates, require further refinement.

With the model calibrated, I conduct three exercises to explore the value of a formal job. In the first exercise, I show that the difference in value functions is on average 13.5 minimum wages, although there is a significant heterogeneity across the firm distribution. Additionally, the gains from holding a formal job decrease as individuals' levels of human capital increase. In the second exercise, I show that receiving a formal job offer at the start of one's career has medium-term consequences for total earnings. Specifically, I find that the lower bound of this effect results in a 6.5% increase in earnings five years later for those who begin their careers in the formal sector.

Next, I analyze the relative importance of the components that distinguish the two sectors. The results show that differences in productivity distribution play a central role in explaining the observed earnings gap between the formal and informal sectors, contributing to nearly a 21% difference in total earnings in the formal sector. This is followed by the variation in the arrival of job offers, which accounts for a 13% difference in total earnings. Additionally, I find that components related to unemployment risk and human capital accumulation are of secondary importance, as their counterfactuals result in smaller decreases in total earnings: 6.3% and 2.9%, respectively.

This paper opens a research agenda that can be expanded in several directions. Using this model and its simulations, it is possible to study the consequences of job loss in the informal sector, an analysis that is not feasible with current data due to the need for long-term observations of workers. Another potential application of this model is to explore the effects of changes in the minimum wage on both informality levels and the value of having a formal job, as discussed in this paper. Currently, the model operates under a partial equilibrium framework and does not account for firm reactions to minimum wage changes. However, it is feasible to incorporate general equilibrium features by assuming a matching, as presented in Flinn (2006).

Finally, as shown in Table 1, there is significant heterogeneity across gender and education, which the current model does not account for. One way to address this is by calibrating the model separately for each group, with the assumption that they participate in distinct labor markets. With these new calibrations, it would be possible to replicate the three main exercises and better understand how the value of a formal job differs for these workers.

References

Abud, M. J., Cases, R., Fuentes, A., & Vergara, R. (2022). *Salario mínimo en Chile: Un aporte a la discusión* (Tech. Rep.). Centro de Estudios Publicos.

Allen, J., Nataraj, S., & Schipper, T. C. (2018). Strict duality and overlapping productivity distributions between formal and informal firms. *Journal of Development Economics*, *135*, 534–554.

Arellano, P., & Jimenez, E. (2016). *Dinámica empresarial. brechas regionales y sectoriales de las pymes en chile (2005-2014)* (Tech. Rep.). Ministerio de Economía, Gobierno de Chile.

Bagger, J., Fontaine, F., Postel-Vinay, F., & Robin, J.-M. (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review*, *104*(6), 1551–1596.

Bobba, M., Flabbi, L., & Levy, S. (2022). Labor market search, informality, and schooling investments. *International Economic Review*, *63*(1), 211-259.

Bobba, M., Flabbi, L., Levy, S., & Tejada, M. (2021). Labor market search, informality, and on-the-job human capital accumulation. *Journal of Econometrics*, *223*(2), 433–453.

Cahuc, P., Postel-Vinay, F., & Robin, J.-M. (2006). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica*, *74*(2), 323–364.

Cases, R., & Fuentes, A. (2024). How substitutable are high-skilled workers? The case of expansion of tertiary education in Chile. *Working paper*.

Cisneros-Acevedo, C. (2022). Unfolding trade effect in two margins of informality. the peruvian case. *The World Bank Economic Review*, *36*(1), 141–170.

Contreras, D., Hurtado, A., & Sara, M. F. (2012). *La excepción chilena y las percepciones de género en la participación laboral femenina* (Tech. Rep.). Departamento de Economía, Universidad de Chile.

Contreras, D., & Plaza, G. (2007). *Participación laboral femenina en Chile.¿Cuánto importan los factores culturales?* (Tech. Rep.). Facultad de Economía y Negocios, Universidad de Chile.

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

Dix-Carneiro, R., Goldberg, P., Meghir, C., & Ulyssea, G. (2021). Trade and informality in the presence of labor market frictions and regulations. *Economic Research Initiatives at Duke (ERID) Working Paper*(302).

Elgin, C., Kose, M. A., Ohnsorge, F., & Yu, S. (2021). Understanding informality. *CAMA Working Paper*.

Encina, J., & Martínez, C. (2009). *Efecto de una mayor cobertura de salas cuna en la participación laboral femenina: Evidencia de Chile* (Tech. Rep.). Facultad de Economía y Negocios, Universidad de Chile.

Flabbi, L., & Tejada, M. (2022). Working and saving informally: The link between labor market informality and financial exclusion. *Working paper*.

Flinn, C. J. (2006). Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates. *Econometrica*, *74*(4), 1013–1062.

Fuentes, A. (2019). *Caracterización del desempleo en Chile* (Tech. Rep.). Centro de Estudios Públicos.

Fuentes, A. (2024). When education is not enough: Explaining labor informality inertia in Chile. *Working paper*.

Haanwinckel, D., & Soares, R. R. (2021). Workforce composition, productivity, and labour regulations in a compensating differentials theory of informality. *The Review of Economic Studies*, *88*(6), 2970–3010.

Hsieh, C.-T., & Olken, B. A. (2014). The missing "missing middle". *Journal of Economic Perspectives*, *28*(3), 89–108.

INE. (2017). *Boletín estadístico: Informalidad laboral. Edición 1* (Tech. Rep.). Instituto Nacional de Estadísticas, Chile.

INE. (2019). *Boletín estadístico: Empleo trimestral. Edición 256* (Tech. Rep.). Instituto Nacional de Estadísticas, Chile.

INE. (2020). *Boletín estadístico: Informalidad laboral. Edición 9* (Tech. Rep.). Instituto Nacional de Estadísticas, Chile.

INE. (2024). *Nota complementaria: Encuesta suplementaria de ingresos 2023* (Tech. Rep.). Instituto Nacional de Estadísticas, Chile.

Jarosch, G. (2023). Searching for job security and the consequences of job loss. *Econometrica*, *91*(3), 903–942.

Jedwab, R., Romer, P., Islam, A. M., & Samaniego, R. (2023). Human capital accumulation at work: Estimates for the world and implications for development. *American Economic Journal: Macroeconomics*, *15*(3), 191–223.

Lagakos, D., Moll, B., Porzio, T., Qian, N., & Schoellman, T. (2018). Life cycle wage growth across countries. *Journal of Political Economy*, *126*(2), 797–849.

La Porta, R., & Shleifer, A. (2008). *The unofficial economy and economic development* (Tech. Rep.). National Bureau of Economic Research.

La Porta, R., & Shleifer, A. (2014). Informality and development. *Journal of Economic Perspectives*, *28*(3), 109–126.

Levy, S., & Schady, N. (2013). Latin america's social policy challenge: Education, social insurance, redistribution. *Journal of Economic Perspectives*, *27*(2), 193–218.

Magnac, T. (1991). Segmented or competitive labor markets. *Econometrica*, *59*(1), 165–187.

Maloney, W. F. (1999). Does informality imply segmentation in urban labor markets? evidence from sectoral transitions in mexico. *The World Bank Economic Review*, *13*(2), 275–302.

Meghir, C., Narita, R., & Robin, J.-M. (2015). Wages and informality in developing countries. *American Economic Review*, *105*(4), 1509–1546.

Parente, R. (2022). *Minimum wages, inequality, and the informal sector.*

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., Mason, A., & Saavedra-Chanduvi, J. (2007). *Informality: Exit and exclusion*. World Bank Publications.

Ruggieri, A., & Cisneros-Acevedo, C. (2023). Firms, policies, informality, and the labor market. *Available at SSRN 4638545*.

Schneider, F., & Enste, D. H. (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature*, *38*(1), 77–114.

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

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

9 Appendix

9.1 Appendix 1: Main job in administrative data

To identify the main job in the administrative records, I follow the procedure outlined in Cases & Fuentes (2024). This procedure involves the following steps: First, identify workers with multiple jobs in the same period, which can be determined because formal workers must pay unemployment insurance for each formal job within a month. For these workers, the primary criterion is to select the job with the highest consistent payment during the overlap period. If multiple jobs have similar payments, the next criterion is to choose the job with the longest tenure duration. Any remaining cases with duplicate entries are dropped, as it is not possible to definitively identify the main job in such instances.

9.2 Appendix 2: Administrative data to panel

To create the panel used in this paper, follow these steps: First, download the 3, 5, and 12% samples of the administrative records from the *Superintendencia de Pensiones* website.[10] The documents consist of a series of .csv files containing data on workers, their individual accounts, requests for unemployment insurance, associated payments, wages, and information on rejected requests. For this paper, only the data on workers and their wages is utilized.

The second step involves merging the workers' data with wage information using the common ID present in both databases. This merge will yield a dataset that includes the worker ID, firm ID, and the wages earned from each job, among other details.

Third, apply the procedure outlined in Appendix 1 to retain only one observation per worker. The remaining data will include information about wages and workers. To compute wage losses due to unemployment, I will create empty observations to indicate unemployment periods between employment spells. I will not add further employment spells beyond the last recorded employment, assuming the worker has retired.

The fourth step involves removing data with potential errors. Specifically, I will drop all observations where the database records a worker as employed but no wage data is available. Additionally, observations with wages below 0.75 minimum wages will be removed, as these are likely indicative of part-time work and to ensure consistency with the procedure outlined in Abud et al. (2022).

[10] The website can be accessed through www.spensiones.cl

Finally, adjust all wages for inflation using the CPI index, covering the period from January 2010 to December 2019.

9.3 Appendix 3: Filters to ensure consistency in tenure

After applying the filters mentioned in the main text, an inconsistency in tenure across survey waves may be observed. This inconsistency arises because there is no process ensuring consistency of tenure reporting between waves; respondents may report more or less tenure on their current job compared to previous waves. To address this issue, I will follow the procedure outlined below.

First, I compute the tenure for each individual in the first wave by calculating the difference in months between the date they report starting the job and the survey date. Next, I compare this tenure with the tenure reported in subsequent waves. If a worker reports being in the same sector in the following wave and indicates a tenure greater than the previous tenure plus three months (since the waves are quarterly), I will adjust their tenure to match the tenure calculated in the first wave plus three months. This procedure will be repeated for each subsequent wave, using the adjusted tenure from the previous wave as the basis. This approach ensures consistency in tenure measures for the specific set of workers.

Second, I will discard several sets of observations associated with workers who have inconsistent labor histories. This includes workers who report being employed in two different firms within the same sector between waves, but where the reported tenure in the first wave exceeds that in the second wave (indicating a firm change), and the tenure reported in the second wave is more than three months, creating an inconsistency. Additionally, I will eliminate observations where there is a sector change (from formal to informal or vice versa) but the reported tenure in the second wave is either higher than in the first wave or exceeds three months.

9.4 Appendix 4: Descriptive statistics administrative data

Table 8 presents descriptive statistics for the formal sector using administrative data. Notably, there are only small differences between the total averages of log wages and tenure in the administrative and survey data. The proportion of women is also roughly consistent with the survey data. However, there is a significant discrepancy in the proportion of workers with tertiary education between the administrative datasets and the surveys. This discrepancy arises because the administrative records capture the level of education workers had at the time they first received formal income. As many workers with tertiary education held formal jobs during their studies, the administrative data tends to underestimate the level of education in the formal sector.

Table 8: Summary statistics - Administrative data (2010-2019)

Year	Average log wages	Average job tenure	Prop. Women	Prop. Tertiary
2010	13.0	25.2	36.0	16.2
2011	13.0	26.6	36.2	16.1
2012	13.1	27.8	36.7	15.9
2013	13.2	29.9	37.5	15.7
2014	13.2	32.4	38.2	15.6
2015	13.2	34.4	38.6	15.3
2016	13.3	36.3	38.9	15.0
2017	13.3	38.2	39.1	14.8
2018	13.3	38.8	39.1	14.5
2019	13.3	40.2	39.2	14.2
Total Average	13.2	33.0	37.9	15.3

9.5 Appendix 6: Algorithm

I solve the model using value function iteration. Specifically, solving the model involves determining the values of firm surpluses and wages for every possible combination of current and benchmark firms. The presence of a minimum wage in the formal sector adds complexity, requiring the computation of the values of the vector $\gamma(\theta, h)$

To solve the model, the surpluses in both sectors must be determined simultaneously. Due to the interdependence between sectors, the minimum wage in the formal sector can influence informal firms, particularly those that are similar to formal firms where the minimum wage is binding. Consequently, it is necessary to estimate the extent to which the minimum wage is binding across formal firms before beginning the solution process. To address this, I propose the following algorithm:

1.1 Guess $\gamma(\theta, h)$

The initial guess I use is 0 for all the firms. Using α is also consistent with this start.

2.1 Guess $S_f(\theta, h)$

The first guess is just a randomly generated matrix with positive values between 0 and 1.

3.1 Guess $S_i(\theta, h)$

The first guess is just a randomly generated matrix with positive values between 0 and 1.

3.2 Compute $U(h)$

This is done using the equation for the value of unemployment.

3.3 $S_i(\theta, h) = \max\{S_i(\theta, h), 0\}$
This is due the definition of the surplus.

3.4 Check and update $S_i(\theta, h)$ until convergence.

2.2 Compute $w(f, \hat{\tau}, \theta, \hat{\theta}, h)$ using $J(\cdot)$
This step implies computing the wage for each feasible combination of current and benchmark firm.

2.3 Update $\gamma(\theta, h) = \max(\max(Aux(f, :, \theta, :, h)))$
Here, I impose $\gamma(\theta, h)$ as the maximum γ observed across all possible benchmarks. The rationale is to associate a firm with the highest γ required to meet the minimum wage, ensuring that the firm can pay it. While it is possible to relax this assumption and consider case-by-case settings, doing so would significantly impact the efficiency of the γ iteration process. Additionally, visual inspection of the different γ values for each firm reveals that they are relatively similar in the different benchmarks..

2.4 $S_f(\theta, h) = \max\{S_f(\theta, h), 0\}$
This is just apply the definition of surplus of a formal firm

2.5 $S_f(\theta, h) = S_f(\theta, h) \times (\gamma(\theta, h) \leq 1)$
Cancel the surplus if the minimum wage is binding.

2.6 Check and update $S_f(\theta, h)$ until convergence.

1.2 Check and update $\gamma(\theta, h)$ until convergence.

Finally, it is important to account for a margin of error in the transitions between firms. Since the surpluses are determined through value function iteration and there is an interconnection between sectors, high precision is required to ensure accurate transitions between firms. To address this, I allow for an additional 0.01 points of surplus as a margin to allow for a job transition.

9.6 Appendix 7: Simulation details

To conduct the simulations, I follow these steps. First, I solve the model using the calibrated parameters. Second, I generate 10 random vectors with dimensions 420 realizations (35 years) each using a uniform distribution. Each vector represents the following components:

1-2 Arrival of formal and informal offers: Take the value 1 if the random realization is below λ_f and λ_i, respectively, and 0 otherwise.

3-6 Arrival of on-the-job offers: Take the value 1 if the random realization is below λ_{ff} and λ_{fi} for the formal sector, and below λ_{if} and λ_{ii} for the informal sector, and 0 otherwise.

7-9 Update of human capital: Take the value 1 if the random realization is below ψ_f, ψ_i and ψ_u, respectively, and 0 otherwise.

10 Firm associated (only for offers): Assign a value between 1 and 25 following a proportional distribution between 0 and 1.

Note that these random vectors are fixed across simulations. This is necessary to prevent variations in the results due to the randomness of the shocks generation.

Once the random realizations are generated, we can begin simulating the model. In the first period, individuals immediately receive offers, as they start unemployed. If they accept an offer, the firm's destruction parameter (θ_δ) is incorporated into their employment history. Also, use the wage function to compute the relevant wage for the worker. Next, another random vector is generated to simulate unemployment shocks, which will occur if the random realization is below θ_δ. Also workers will incorporate unemployment as the benchmark.

Once workers begin receiving on-the-job offers, they will initially accept any firm as their benchmark, as their current benchmark value is unemployment, equivalent to zero. After updating the benchmark, they will use the wage function to compute the new relevant wage. From that point on, whenever they receive an on-the-job offer, they will compare the surplus of the new offer with that of their current firm and the benchmark to decide their next move. Each time a worker switches to a new firm, their previous firm will then serve as the updated benchmark.

If workers experience an unemployment shock, they will exit both their current firm and benchmark, entering the unemployment pool and potentially facing human capital de-accumulation shocks.

Also, workers can increase their levels of human capital if they survive the destruction realization.

After completing these steps, the labor history of one worker is generated. To simulate the entire economy, I repeat this procedure for 10,000 workers, creating a comprehensive simulated dataset.

9.7 Appendix 8: Simulation for top firms

Table 9 presents the simulation results for the top firms in each joint distribution. This scenario is relatively rare, with the probability of receiving an offer from the top firm in the formal sector being only 0.38% ($\lambda_f \times \frac{1}{25}$). The results indicate a significant difference in earnings in the first year, with the formal firm paying more than twice what the informal firm pays. The disparity in total earnings is largely driven by differences in productivity, as the unemployment periods in

both cases are minimal.

As with the previous exercises, the differences in earnings tend to diminish over time. By the third year, the earnings disparity has narrowed to about 75%, primarily due to a reduction in the average wage differences. By the fifth year, the differences have decreased but remain significant, accounting for nearly 42%.

Table 9: Labor outcomes from simulation after 1,3 and 5 years for highest firm.

	1 Year - Highest		
	Formal	Informal	Difference
Earnings	22.92	10.25	123.7%
Average wage	1.96	0.89	119.7%
Unemployment periods	0.32	0.45	-30.0%

	3 Years - Highest		
	Formal	Informal	Difference
Earnings	63.08	36.06	74.9%
Average wage	1.83	1.08	69.9%
Unemployment periods	1.71	2.46	-30.8%

	5 Years - Highest		
	Formal	Informal	Difference
Earnings	98.78	69.68	41.8%
Average wage	1.75	1.27	38.1%
Unemployment periods	3.61	4.94	-26.9%