

Assessing the Impact of Informal Sector Employment on Young Less-Educated Workers*

Javier Cano-Urbina[†]

John Gibson[‡]

Florida State University

University of Texas at El Paso

November 8, 2023

Abstract

In this paper, we develop a search and matching model that allows for two important channels through which participation in the informal sector may benefit young less-educated workers: (i) human capital accumulation, and (ii) employer screening. We calibrate our model using the ENOE, a Mexican household survey on income and labor dynamics. Using our calibrated model, we shed light on many unobservable characteristics of the Mexican labor market for young less-educated workers, most notably the differing hiring standards for informal and formal jobs. Specifically, hiring standards for these workers are found to be substantially higher for formal versus informal positions, making these workers naturally flow from unemployment to the informal sector where they can gain skills and reduce the uncertainty about their abilities. We also conduct counterfactual policy experiments to assess how labor market reforms designed to limit the employment share of the informal sector impact young less-educated workers. While our results favor reducing regulations in the formal sector due to the policy's positive effect on aggregate employment, the policy is still found to impede the development of skills within this vulnerable segment of the labor market.

Keywords: Informal Sector, Labor Search, Employer Screening, Human Capital Formation.

JEL Classification: J46, J64, J68, O17, E24, E26

*We would like to thank Santanu Chatterjee, Lei Fang, Lance Lochner, Federico Mandelman, Pierre Nguimkeu, Felix Rioja, and Jeffrey Smith for very helpful comments as well as seminar participants at Tulane University, Banco de México, El Colegio de México, and the conference participants at the Southern Economic Association Meetings, Southeastern International/Development Economics Workshop, International Computing in Economics and Finance Conference, the Canadian Economic Association Meetings, and the Midwest Macroeconomic Meetings

[†]Department of Economics, Florida State University, Tallahassee, FL 32306, USA. Telephone: 850-644-7090. Email: jcanourbina@fsu.edu

[‡]Department of Economics and Finance, University of Texas at El Paso, El Paso, TX 79968, USA. Telephone: 915-747-5435. Email: jdgibson@utep.edu

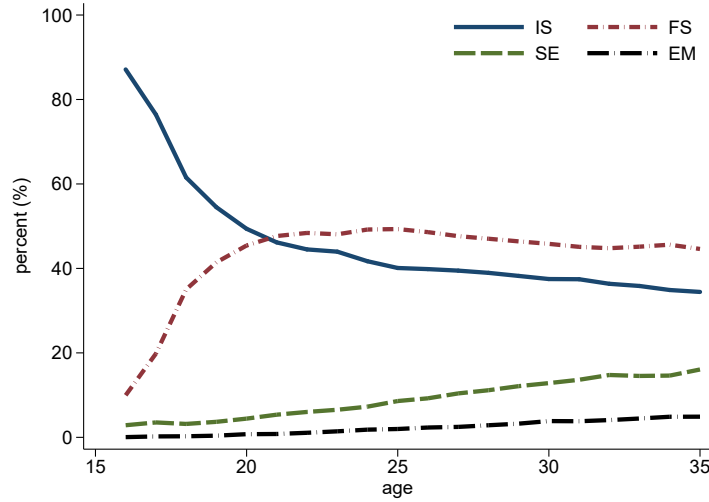
1 Introduction

In developing countries a substantial fraction of workers are employed through irregular *under-the-table* jobs. Jobs such as these are typically referred to as “informal jobs,” and taken together, constitute the informal sector. Traditionally, the informal sector carries negative connotations such as poor working conditions, low pay, lack of basic benefits, and lower productivity. It has been argued that the existence of a large informal sector can negatively affect growth by congesting public services (e.g., Loayza, 1996). In a more recent study, Horvath and Yang (2022) highlight both a cost and a benefit associated with large informal labor markets, arguing that a large informal sector amplifies the impact of productivity shocks on output while reducing their impact on employment. Finally, other studies consider the possible benefits that the informal sector may provide to specific sub-groups within the labor market, arguing that jobs in the informal sector provide young low-skilled workers with employment and worker-training opportunities outside of those offered in the formal sector or by the government (Hemmer and Mannel, 1989).

In the spirit of this last argument, our study examines the extent to which informal sector employment provides young less-educated workers with opportunities to advance their careers and transition into formal-sector jobs. Our focus is motivated by the labor market experience of young less-educated workers in Mexico. Figure 1 describes the distribution of less-educated workers by sector of employment and age using the ENOE (*Encuesta Nacional de Ocupación y Empleo*), a detailed household survey from Mexico. Inspection of Figure 1 clearly shows that the youngest less-educated workers are predominantly employed in the informal sector. Furthermore, moving across age cohorts, we find that the proportion of workers employed in the informal (formal) sector decreases (increases) monotonically with age. This pattern of employment adjustment continues until workers reach their mid-twenties, at which point sectoral employment shares remain more or less constant. This pattern is consistent with the idea that informal-sector jobs can serve as a “port-of-entry” into the labor market for young less-educated workers (Arias and Maloney, 2007), and it supports the notion that informal positions have lower barriers to entry than their formal-sector counterparts (Fields, 2009). If these are true, and if the formal sector offers better job opportunities than the informal sector, then the distribution of less-educated workers in the formal and informal sector presented in Figure 1 is consistent with the possibility that informal-sector jobs offer less-educated workers some opportunities that facilitate their entry into the formal sector.¹

¹Note that Figure 1 makes clear that we are treating informal-sector workers separately from self-employed workers. That is, when we refer to the informal sector, we refer to those salaried workers. Some studies pool informal salaried and self-employed in the same category. Our reason to separate informal salaried from self-employed corresponds to our interest in human capital accumulation and screening of workers skills,

Figure 1: Distribution of Less-Educated Workers by Employment Sector and Age in Mexico



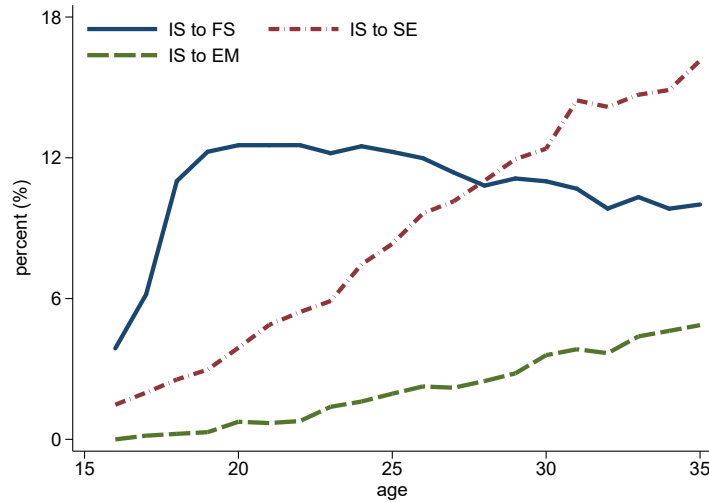
Notes: Male workers with less than 12 years of education. The lines correspond to the proportion of workers employed in each sector by age. In the graph IS corresponds to informal-salaried workers, FE to formal-salaried, SE to self-employed, and EM to individuals that own their own firm and become employer or boss of salaried workers. Data obtained from the Mexican survey ENOE (*Encuesta Nacional de Ocupación y Empleo*) from the first quarter of 2005 to the fourth quarter of 2018.

We focus on two possible channels through which informal-sector jobs may facilitate entry into the formal sector for less-educated workers: (i) informality may provide workers with the opportunity to accumulate human capital and gain skills that are valued by formal-sector employers, and (ii) informality may provide firms with a cost effective way to screen employees and reduce the uncertainty about the worker's skill level. Both, the lack of skills and the uncertainty about the worker's skill level could be barriers to entry into formal jobs for less-educated workers.

The existing literature provides empirical evidence for each of these mechanisms. Cano-Urbina (2016) finds evidence that is consistent with informal jobs offering opportunities for general human capital accumulation for less-educated workers in Mexico but this study does not consider the role of screening. Cano-Urbina (2015) develops two parallel models: (i) one model that considers human capital accumulation but ignores screening, (ii) another model that considers screening but ignores human capital accumulation. This latter study finds evidence that is consistent with the analytical implications of the model that includes screening but not human capital accumulation. In all likelihood, the informal sector plays both roles in the career development of less-educated workers, and so the current study extends

which might not be very relevant, specially screening, for a self-employed worker.

Figure 2: Transition Rate of Less-Educated Workers in Mexico by Age



Notes: Male workers with less than 12 years of education. Number of transitions as a fraction of the number of workers in the informal sector. In the graph IS corresponds to informal-salaried workers, FE to formal-salaried, SE to self-employed, and EM to individuals that own their own firm and become employer or boss of salaried workers. Data obtained from the Mexican survey ENOE (*Encuesta Nacional de Ocupación y Empleo*) from the first quarter of 2005 to the fourth quarter of 2018.

Cano-Urbina (2015) by developing a model that considers both mechanisms simultaneously within a single theoretical framework. Since the theoretical model is complicated, deriving analytical implications is no longer an option, and so we proceed in the current study by exploring these two mechanisms computationally.

The importance of introducing both mechanisms within a single model environment can be seen in Figure 2. Notice that the adjustment from informal to formal employment occurs at a non-linear rate, with rapid transitions occurring for the first few years, say from ages 16 to 19, followed by more gradual transitions over the next several years. This pattern of adjustment is consistent with rapid employer screening during individuals' first few working years, followed by a gradual rate of human capital accumulation. This suggests that both mechanisms are needed to account for the dynamics observed in the data.²

We use data from the ENOE to estimate the rates of employer screening and human capital accumulation for both the formal and informal sectors. The estimation is based on the model's predictions regarding wage changes. In particular, we use the model's predictions

²The way we model both mechanisms in this paper is very general and can account for other more elaborate or specific mechanisms that give an important role to informal jobs in the early careers of less-educated workers. It becomes clear below that the employer screening mechanism also involves workers learning about their skills (and interests). These mechanisms can also interact so that human capital accumulation can work both as a screening mechanism and to help some workers accumulate skills.

to identify wage changes that can be attributed to employer learning (screening) or human capital accumulation. We find that informal sector jobs have higher rates of both employer learning and human capital accumulation than formal sector jobs. Our estimates suggest that it takes on average 24 months for an employer in the formal sector to determine the skill level of a new labor market entrant, whereas this same process takes on average 18 months in the informal sector. Similarly, our estimates suggest that it takes on average 36 months for a low skilled worker to accumulate skills while employed in the formal sector and it takes on average 28 months if the low skilled worker is employed in the informal sector.

At first glance, the previous findings seem to contradict Bobba et al. (2019), who finds that the formal sector, rather than the informal sector, offers a higher rate of human capital accumulation. However, the faster human capital accumulation in the formal sector “is partially offset by the size of the upgrade when the shock hits, ..., the average size of the jump is larger when working informally” (see page 21 in Bobba et al., 2019). That is, the increase in the level of human capital is larger when working informally. In our model the “jump” when accumulating skills is of the same magnitude in the formal and informal sectors. As a result it is not clear that their results contradict our findings³.

Recent evidence provided by Engbom (2022) also supports our finding that informal jobs are more conducive of skill accumulation than formal jobs. In particular, Engbom (2022) develops a life-cycle model of the accumulation of skills in a frictional labor market and finds that a more fluid labor market facilitates higher-quality matches in terms of skills, thus incentivizing the accumulation of additional skills. To the extent that the informal labor market for less-educated workers in Mexico is more fluid than its formal counterpart, this result suggests that less-educated workers should accumulate skills faster in the informal sector than in the formal sector. The results from Engbom (2022) are also consistent with the findings of Cano-Urbina (2016) which finds that less-educated workers in the informal sector experience faster wage growth than less-educated workers in the formal sector, suggesting that human capital accumulation is faster in the informal sector for this group of workers.

Our results suggest that those workers who are revealed as more productive while employed in the informal sector increase their chances of finding a formal job. This is consistent

³It is also possible that the skills in Bobba et al. (2019) differ in nature from the skills in our model. In Bobba et al. (2019), skills previously accumulated on-the-job could be lost when the worker is unemployed but not while the worker is continuously employed by the same firm. Since skills can only be lost while changing firms through an unemployment spell then some of these skills might not be fully portable from one job to another. In terms of the classical human capital theory of Becker (1993), this means that skills captured in Bobba et al. (2019) are to some extent firm- or industry-specific. In our model, skills are fully portable from one job to another, so that they are general skills. This can also explain why human capital accumulation rates are larger in the formal sector than in the informal sector in Bobba et al. (2019). If formal firms have more firm- or industry-specific technology than informal firms, there will be faster firm- or industry-specific skill accumulation in the formal sector than in the informal sector.

with the findings of Samaniego de la Parra and Fernandez Bujanda (2020). They explore how increased monitoring of labor regulations by authorities affect a series of outcomes such as job creation and job destruction. The authors find that increasing monitoring both increases job destruction in the informal sector and facilitates transitions into the formal sector. If an informal worker is formalized as a result of higher monitoring, then employers may take this as a signal of the worker’s productivity, increasing the likelihood that they are poached by a better formal job.

In this paper, we introduce both employer screening and human capital accumulation within a two-sector labor search model. In our model, firms create vacancies in either the formal or the informal sector. Positions in the formal sector benefit from higher average productivity and, as a consequence, offer higher wages on average. However, firms operating formally are required to pay firing costs whenever a worker separates from the firm. While there are certainly other regulatory burdens associated with the formal sector (e.g., taxation, costly worker benefits, etc.), we abstract from these costs and focus attention on the important role played by firing costs. Positions created informally are not subject to firing costs, but are on average less productive and generally offer lower wages. Also, there is a central authority (or government) that monitors informal activity. If a firm is caught operating informally, its position is destroyed and the firm faces a fine or penalty.

Workers in our model differ in terms of their skill level, which may be high or low. While agents are endowed with a specific skill level at birth, we assume that all workers enter the labor market with an unknown/unobserved skill level that is revealed while working in either sector. Therefore, workers in our model exist in one of three states related to their skills: unknown, high, or low. If a worker is revealed as being low-skilled, they can become high-skilled over time by working in either sector. Once a worker becomes high-skilled, they remain high-skilled until they permanently exit the labor market. The processes by which workers’ skill levels are revealed and low-skilled workers transition to high-skilled correspond to the employer screening and human capital accumulation mechanisms, respectively.

Workers and firms meet through a standard matching process where unemployed workers search for employment randomly across both sectors, regardless of their skill level. Once a worker and firm meet, they draw a match-specific productivity from a known distribution and only meetings with match-specific productivity larger than a hiring standard result in the creation of a job. If the match-specific productivity is lower than the hiring standard, both workers and firms keep searching for better matches. As a consequence, endogenous hiring cutoff rules will be formed to determine the minimum match-specific productivity needed before forming a productive match (e.g., hiring the worker). Given that workers may exist in one of three skill states (unknown, high, or low) and that employment may occur in

one of two sectors (formal and informal), our model will generate six different endogenous hiring cutoff rules. Variation across these cutoff rules in our baseline model can be used to provide evidence in support of the “port-of-entry” role for informal employment. Finally, once workers find a job in the informal sector, we assume that they keep searching for formal employment. This allows the model to generate direct transitions from informal to formal employment, which are found to be very prevalent in the data.⁴

The results of our calibrated model provide evidence in support of the “port-of-entry” role of informal employment. Specifically, we find that the endogenous hiring cutoffs for agents with unknown skill level differ substantially across sectors, with the informal sector cutoff being significantly less restrictive than the cutoff for the formal sector. As a consequence, our model suggests that new labor market entrants are more likely to flow from unemployment into temporary informal-sector employment until they have their skill-type revealed and gain human capital. Furthermore, the calibration reveals the importance of both screening and human capital accumulation. If we start with an economy in which all workers enter the labor market with their skill level unknown, in the steady-state equilibrium the economy only has about 13% of workers with an unknown skill level, consistent with substantial employer screening. Similarly, if we start with an economy in which nearly half of the workers are low-skilled and the other half are high-skilled, in the steady-state equilibrium, 91.5% of the workers with a known skill level are high-skilled, consistent with substantial skill accumulation.⁵ If we add to these findings the fact that most of the workers with low or unknown skill level are employed in the informal sector, it highlights the role that informal jobs have in the careers of these workers.

Equipped with the calibrated model, we investigate two alternative policies that both intend to reduce the number of informal-sector jobs in the economy. Such policies typically focus on firm behavior by either (i) increasing the punishment for non-compliance among informal firms, or (ii) reducing the regulatory burden associated with formal-sector participation. While it is not surprising that the counterfactual experiments reveal that reducing the regulatory burden results in lower unemployment than increasing punishment for non-compliance, there are other interesting findings related to these policies in regards to the screening and human capital accumulation processes. Both policies boost the creation of formal-sector vacancies and reduce the creation of informal-sector vacancies. A second effect

⁴This type of on-the-job search is necessary for our model to capture the large rate of transition between informal and formal employment. While there is also empirical justification for allowing some level of direct transitions from formal to informal employment, we abstract from this aspect of on-the-job search in order to focus on understanding the role of the informal sector in the career development of less-educated workers.

⁵As described further below, the proportion of low- and high-skilled workers is determined using household survey data. However, the estimate comes out to roughly a 50/50 split between high and low skilled.

of these policies is to move the hiring standards in both sectors. While the formal-sector hiring standards decrease with both policies, this decrease is not enough to make them lower than the informal-sector hiring standards. As such, the informal sector remains a port-of-entry for many workers, and reducing informal vacancy creation reduces both employer screening and human capital accumulation. Both policies are found to reduce the share of high-skilled workers and increase the shares of low-skilled workers and workers with unknown skills, though the distortions created by increasing punishment are found to be larger. As such, policies designed to reduce informal employment should be complemented by additional policies designed to help formal-sector jobs perform the roles of employer screening and human capital accumulation for jobs held by less-educated workers⁶.

The rest of the paper is organized as follows. Section 2 presents the mechanisms of our model, while Section 3 characterizes the equilibrium. Section 4 describes the survey data we use for calibration and Sections 5 and 6 detail our estimation and calibration strategies, respectively. Section 7 presents our results and Section 8 concludes.

2 Theoretical Framework

Our theoretical framework is motivated by empirical observations from Mexico. Figure 1 suggest that young less-educated workers in Mexico have a much harder time finding formal-sector jobs than informal-sector jobs. However, this pattern breaks down quickly as workers age, suggesting that the additional experience or skills accumulated in the informal sector help facilitate transitions into formal employment. As we develop our model below, we will keep coming back to these facts.

⁶Our policy experiments are also related to the literature on the relationship between job security regulations, worker protection, and employment outcomes. Existing works in this area have reached different conclusions regarding the impact of such regulations. For example, while Card and Krueger (2000) reports minimal impact on employment following an increase in the minimum wage, Heckman et al. (2000), finds that “job security policies have a substantial impact on the level and distribution of employment in Latin America.” Specifically, insiders gain at the expense of outsiders, suggesting that job security provisions may disproportionately impact young less-educated workers as they enter the labor market. Similarly, Montenegro et al. (2007) find that firing cost disproportionately impact young workers in Chile, as firing costs rise with worker tenure thereby discouraging employers from firing more senior workers during times of economic hardship. In our model, firing costs are the only job security provision. Our experiments suggest that lowering firing costs will reduce the size of the informal sector, lower unemployment, and results in fewer high-skilled workers in equilibrium. While these findings support the notion that lowering job security provisions expands the formal labor market and increases employment opportunities for workers, the policy also carries the unintended consequence of weakening the skill revelation and human capital accumulation mechanisms present within this segment of the labor market.

2.1 Employment Sectors

We consider a labor market with two sectors, one formal and one informal. Firms in the formal sector are on average more productive than firms in the informal sector. This difference in average productivity between sectors could be a result of formal-sector firms having access to better outside financing and more investment in physical capital than informal-sector firms (as modeled in Amaral and Quintin, 2006). Formal and informal firms face different institutional frameworks: (i) formal-sector firms are subject to a firing cost incurred when matches are destroyed whereas informal-sector firms do not incur this cost, and (ii) informal-sector firms are subject to a penalty if they are caught by the authorities in which case the job is destroyed whereas formal-sector firms do not incur this penalty.

When modeling the regulatory environment of the formal sector, we focus all attention on the costs associated with firing, and abstract away from other regulatory costs associated with formal employment, such as paid benefits and payroll taxes. This choice allows us to focus on the aspect of formal-sector regulation that results in differential hiring decisions across skill level, while abstracting away from regulations that impact workers symmetrically. Specifically, firing costs are only paid if a worker separates from a firm, while paid benefits and payroll taxes must be paid for the duration of the employment relationship. As such, firing costs have the potential to distort hiring decisions on the basis of expected termination risk. This may make employers more reluctant to hire workers of unknown skill level due to concerns that they may have to be dismissed in the near future. Such a distortion is not present with other regulatory costs. Given that we are interested in our model's endogenous hiring rules for workers with both known and unknown skill levels, it is important that we account for empirically reasonable firing costs.

It is also important to note that firing costs are very large in Mexico and represent a sizable component of the regulatory burden placed on formal sector firms. In a cross country study of regulatory costs in Latin America, Heckman et al. (2000) find that Mexico consistently ranks above average in terms of firing costs, suggesting a high degree of regulation related to job security and retention. Montes Rojas and Santamaría (2007) also stress that the mandated fees imposed when terminating a worker may only represent a fraction of the true cost of the separation. Specifically, firms are required to cover legal fees and pay workers' forgone wages during court deliberations when terminations are contested. As such, while the authors report the average direct cost of firing a worker in Mexico is approximately six times quarterly wages, they state that once legal fees and back pay are taken into account, this figure can increase by upwards of 50%. This point is discussed further in Section 6 when we calibrate the model.

2.2 The Search and Matching Process

We assume that workers enter the labor market with different skill levels. While we can conceive of a continuum of skill levels, we simplify the analysis by considering only two skill levels: low or high. However, when workers first enter the labor market their skill level is unknown. We assume that neither the worker nor the firm know the worker's skill level, so that information about the new worker's skill level is symmetric. All that is known is that a fraction, ν , of new workers are low-skilled (L-skilled) and a fraction, $1 - \nu$, are high-skilled (H-skilled). We refer to new workers with unknown skill level as "newcomers." All newcomers enter the labor market through unemployment where they search for jobs in both the formal and informal sectors. When newcomers find a job in the formal or informal sector their skill level can be revealed. Skills cannot be revealed while the newcomer is unemployed.

All unemployed workers, search for jobs in both the formal and informal sectors, irrespective of their skill level. Given this job-search behavior, unemployed workers contact firms in the informal sector according to the function:

$$(1) \quad m_I(u, v_I) = \gamma_1^I u^{\gamma_2^I} v_I^{1-\gamma_2^I}$$

where u is the total unemployment rate across all skill types, and v_I is the number of vacant informal-sector jobs as a fraction of the labor force. The total unemployment rate is defined as $u = u_N + u_H + u_L$, where u_N , u_H , and u_L are the unemployment rates of newcomers, H-skilled, and L-skilled workers, respectively. As a result, informal-sector firms contact unemployed workers at rate $q_I = m_I(u, v_I)/v_I = \gamma_1^I \theta_I^{-\gamma_2^I}$, where $\theta_I = v_I/u$ is the labor-market tightness in the informal sector. Similarly, unemployed workers contact informal-sector firms at rate $f_I = m_I(u, v_I)/u = \gamma_1^I \theta_I^{1-\gamma_2^I}$.

When a worker finds a job in the informal sector, we assume that this worker keeps searching for a job in the formal sector and that they move to a formal-sector job whenever it is optimal for them to do so.⁷ Given this job-search behavior, unemployed and informal-sector workers contact formal-sector firms according to the following function:

$$(2) \quad m_F(u + n_I, v_F) = \gamma_1^F (u + n_I)^{\gamma_2^F} v_F^{1-\gamma_2^F}$$

where n_I is the total number of workers employed in the informal sector across all skill types, as a fraction of the labor force and v_F is the number of vacant formal-sector jobs as a fraction of the labor force. Following Bosch and Esteban-Pretel (2012) we allow for the

⁷To simplify analysis, and to focus attention on direct transitions from the informal sector to the formal sector, we abstract away from all other forms of on-the-job search.

possibility that unemployed and informal-sector workers are not equally efficient in their job search efforts. As such, while formal-sector firms contact unemployed individuals at the rate $q_F = m_F(u + n_I, v_F)/v_F = \gamma_1^F \theta_F^{-\gamma_2^F}$, informal-sector workers are contacted at the rate ηq_F .⁸ In the contact rates described above, $\theta_F = v_F/(u + n_I)$ denotes the labor market tightness in the formal sector and η is the parameter that summarizes the relative efficiency of job search behavior between unemployed and informal-sector workers. Similarly, unemployed individuals contact formal-sector firms at the rate $f_F = m_F(u + n_I, v_F)/(u + n_I) = \gamma_1^F \theta_F^{1-\gamma_2^F}$ while informal-sector workers contact these firms at the rate ηf_F .

Figure 1 suggest that less-educated workers face significant barriers to access formal-sector jobs and much lower barriers to access informal-sector jobs. As such, we focus attention on the endogenous hiring process and assume that not all contacts between job seekers and firms result in a match being created. Specifically, we assume that when a firm and job seeker meet, they must draw a match-specific quality, denoted by x , from a known distribution, $G(x)$, defined over $[0, 1]$. For the worker-firm pair to begin producing (e.g. for the job seeker to be hired), their match quality must exceed the endogenous hiring standard for their segment of the labor market.⁹ If a productive match is successfully formed, the match quality remains constant until the match is destroyed. As we show below, hiring standards will depend on both the sector of employment (formal or informal) and the worker's skill level at the time of contact (newcomer, H-skill, or L-skill). We denote the hiring standards in sector $j \in \{F, I\}$ as C_{jN} , C_{jH} , and C_{jL} for newcomers, H-skilled, and L-skilled workers, respectively, where $j = F$ corresponds to the formal sector and $j = I$ corresponds to the informal sector. We do not impose any ordering on these hiring standards across skill-type or sector of employment, but instead determine their values as an equilibrium outcome.

A key feature of our model is that the vacancies posted by firms are not skill specific. That is, v_I and v_F are the number of total vacancies posted in each sector, and these vacancies can be filled with a newcomer, H-skilled, or L-skilled worker. The rationale for this set up is that the empirical motivation of our study is based in the group of less-educated workers, which implies that: (i) skills are not tied to educational attainment, (ii) the type of jobs that these workers access are less likely to require the manipulation of sophisticated equipment, and (iii) part of the skill accumulation also involves work ethics that more educated workers could accumulate in the formal educational system. Instead, we assume that if a vacancy

⁸Bosch and Esteban-Pretel assume that $\eta < 1$ so that informal-sector workers are less efficient in their search efforts than unemployed workers (see first paragraph of page 277 of Bosch and Esteban-Pretel, 2012). Faberman et al. (2017) suggests that on-the-job search in the US labor market is more efficient than search while unemployed which would imply that $\eta > 1$. We follow Bosch and Esteban-Pretel in our calibration exercise below as we believe this is a more realistic assumption within the context of informal labor markets.

⁹The interested reader is directed to Chapter 6 of Pissarides (2000) for a textbook treatment of the mechanism described above.

is contacted by a L-skilled worker, the firm will only hire the worker if the match quality is sufficiently high (i.e., larger than C_{jL} , $j \in \{F, I\}$). Therefore, a firm does not have a preference for H-skilled workers *per se*, since a L-skilled worker with a very high match quality could produce more than a H-skilled worker with a smaller match quality.¹⁰ This point becomes clear when we define the firms' value functions below.

We assume that both formal and informal-sector matches are exogenously destroyed at fixed Poisson rates, where λ_F and λ_I are the rates in the formal and informal sectors, respectively. However, job destruction is not purely exogenous in our model as jobs can be endogenously destroyed when a newcomer's skill-level is revealed. That is, suppose that the hiring standards in the formal sector are such that $C_{FN} < C_{FL}$ and that the current match quality between a newcomer and a formal-sector firm is $\tilde{x}$, where $C_{FN} < \tilde{x} < C_{FL}$. If the newcomer is revealed to be L-skilled this formal-sector match will be destroyed. The same argument applies to an informal-sector firm. In fact, endogenous job destruction due to skill revelation could even happen for H-skilled. It all depends on the ordering of the endogenously determined hiring standards, C_{jk} , in each sector. Another endogenous job destruction mechanism in our model takes place in the informal sector when informal-sector workers contact a formal-sector firm with an open vacancy and draw a match quality $x > C_{Fk}$, $k \in \{N, H, L\}$ leaving the informal-sector firm.

Crucially, both forms of job destruction in the formal sector require firms to pay firing costs, $D > 0$. Firms that operate informally avoid paying firing costs but are subject to monitoring by authorities. If caught, which occurs with probability π , the informal-sector job is destroyed and the firm must pay a fine of $T \geq 0$. Finally, we assume that all workers exit the labor market permanently at an exogenous rate, τ . This form of job destruction does not result in firing costs for firms in the formal sector. Every worker who exits the labor market through this channel is replaced by a newcomer who enters the labor market through unemployment. A fraction, ν of these newcomers will be L-skilled while the remaining fraction, $1 - \nu$, will be H-skilled.

2.3 Skill Revelation

While all workers enter the model as newcomers with an unknown skill level, they can have their skill level revealed while employed in either sector. This skill revelation process could be modeled in a way similar to Jovanovic (1979) so that the inclusion of noise in the production process causes a firm and a newcomer to gradually learn the newcomer's skill

¹⁰A similar argument can be applied for newcomers. That is, a newcomer with a very high match quality could produce more than a H-skilled worker with a low match quality.

level.¹¹ This process could then be calibrated so that a newcomer's skill level is expected to be revealed within S periods and that the corresponding average rate of employer learning is $1/S$ per period. Given that our interest lies not in the specific learning model, but rather how employer learning affects the dynamics of hiring and firing in an economy with a large informal sector, we can substantially simplify the process by defining exogenous rates of employer learning. Specifically, we define $\sigma = 1/S$ as the average rate of employer learning and then take σ as the steady-state probability that a newcomer's skill level is learned or revealed. Therefore, we assume that employer learning is a stochastic process such that at every moment a newcomer's skill level can be revealed with probability σ_F if employed in the formal sector and σ_I if employed in the informal sector.

Along with the assumption that a newcomer's skill level is revealed with probability $1/S$ per period, we also assume that newcomers produce an intermediate level of output, p_{jN} , which lies between p_{jL} and p_{jH} , for $j \in \{F, I\}$. While this is technically not true, as newcomers' true skill level is simply unknown, it simplifies our model while focusing on the mechanism of primary significance, the origination of the labor contract. Assuming that newcomers possess intermediate skills is similar to firms' making the hiring decision based on the newcomers' expected productivity. The only difference is that assuming that newcomers produce an intermediate level of output will result in instances where worker-firm pairs realize profits (if the newcomer is actually H-skilled) or losses (if the newcomer is actually L-skilled). Given that data suggests that newcomers are nearly evenly distributed between H-skilled and L-skilled, these profits and losses would cancel out in aggregate.¹² Furthermore, given that our focus is on the flow of workers in the labor market, the profits and losses that may originate as a result of these hires is not of central importance.

To illustrate the importance of the skill revelation mechanism in our model environment, consider the value of a newcomer to a firm in either employment sector. Specifically, let $J_{jk}(x)$, $j \in \{F, I\}$, $k \in \{N, H, L\}$, be the present discounted value of expected profit from a job in sector j occupied by a worker of skill k and match quality x . Similarly, let V_j be the present discounted value of expected profit from a vacant job in sector j . An informal-sector

¹¹That is, if p_l is the true productivity for a worker of skill level l , we can think that the newcomer's observed output is $y = p_l + \varepsilon$ where $\varepsilon \sim \Psi(\varepsilon)$. Then every period, the firm and newcomer observe y and update their belief about the value of p_l . After some period of time both the newcomer and the firm eventually learn the true value p_l . Given $\Psi(\varepsilon)$, we could find the average number of periods, S , that it takes to learn the worker's true skill level.

¹²In Section 6 and Appendix E.4 we describe the procedure to determine the fraction of L-skilled entering the labor market, ν , and the results suggest that $\nu \approx 0.5$.

firm matched with a newcomer with match quality x faces the following value function:

$$(3) \quad rJ_{IN}(x) = p_{IN}x - w_{IN}(x) + (\lambda_I + \eta f_F \bar{G}(C_{FN}) + \tau)[V_I - J_{IN}(x)] + \pi[V_I - T - J_{IN}(x)] \\ + \sigma_I \nu \left(\Gamma_{IL}(x)[J_{IL}(x) - J_{IN}(x)] + (1 - \Gamma_{IL}(x))[V_I - J_{IN}(x)] \right) \\ + \sigma_I(1 - \nu) \left(\Gamma_{IH}(x)[J_{IH}(x) - J_{IN}(x)] + (1 - \Gamma_{IH}(x))[V_I - J_{IN}(x)] \right)$$

where p_{IN} is the productivity of a newcomer employed in an informal-sector firm, so the output from this match is $p_{IN}x$ and the firm pays the newcomer wage $w_{IN}(x)$ that also depends on the match quality. The value function indicates five reasons for termination of informal-sector jobs. The first three are: (i) exogenous job destruction, (ii) the newcomer finds a formal-sector job and quits, and (iii) the newcomer permanently exits the labor market. In these three cases, the firm has a loss of $[V_I(x) - J_{IN}(x)]$. The fourth reason an informal-sector job separates is that the employment relationship is detected by the authorities and terminated. In this case, the firm faces both a job separation as well as the penalty T and has a loss of $[V_I(x) - T - J_{IN}(x)]$. The fifth reason involves the skill revelation process. With probability σ_I the newcomer's skill level is revealed. If the worker's skill level is revealed, there is a probability ν that the worker will be L-skilled and a probability $(1 - \nu)$ that the worker will be H-skilled. Then, regardless of the skill level, the job can be destroyed if the match quality x is not high enough for the worker's skill level, which is represented by the indicator functions $\Gamma_{Ik}(x) = \mathbf{1}\{x > C_{Ik}\}$ for $k \in \{H, L\}$. Notice that this type of job destruction does not generate firing costs because informal-sector firms avoid this cost so the firm experiences a loss of $[V_I(x) - J_{IN}(x)]$. If match quality is large enough to continue the productive relationship after the skill level is revealed, the firm and the worker renegotiate the labor contract and the firm experiences a gain of $[J_{Ik}(x) - J_{IN}(x)]$, for $k \in \{H, L\}$. The value function $J_{Ik}(x)$ is described in detail in the next section.

Similarly, a formal-sector firm matched with a newcomer with match quality x faces the following value function:

$$(4) \quad rJ_{FN}(x) = p_{FN}x - w_{FN}(x) + \lambda_F[V_F - D - J_{FN}(x)] \\ + \sigma_F \nu \left(\Gamma_{FL}(x)[J_{FL}(x) - J_{FN}(x)] + (1 - \Gamma_{FL}(x))[V_F - D - J_{FN}(x)] \right) \\ + \sigma_F(1 - \nu) \left(\Gamma_{FH}(x)[J_{FH}(x) - J_{FN}(x)] + (1 - \Gamma_{FH}(x))[V_F - D - J_{FN}(x)] \right) \\ + \tau[V_F - J_{FN}(x)]$$

where p_{FN} is the productivity of a newcomer employed in a formal-sector firm and the firm pays the newcomer wage $w_{FN}(x)$. The value function indicates three reasons why formal-sector jobs can be destroyed: (i) exogenous job destruction, (ii) newcomer permanently exits the labor market, and (iii) the newcomer skill level is revealed and match quality is not high enough for the revealed type. If the worker permanently exits the labor market the firm does not incur firing costs and the firm has a loss of $[V_F - J_{FN}(x)]$. But if the job is terminated exogenously or after skill revelation the firm incurs firing costs D resulting in a loss of $[V_F - D - J_{FN}(x)]$. Job termination after skill revelation occurs if the match quality is not high enough represented by function $\Gamma_{Fk}(x) = \mathbf{1}\{x > C_{Fk}\}$ for $k \in \{H, L\}$. As in the case of the informal sector, if the match quality is high enough, the productive relationship continues, the labor contract is renegotiated, and the firm experiences a gain of $[J_{Fk}(x) - J_{FN}(x)]$, for $k \in \{H, L\}$. The value function $J_{Fk}(x)$ is described in detail in the next section.

2.4 Skill Accumulation

Once a worker's skill level is revealed, those workers who are found to be L-skilled can accumulate human capital and become H-skilled while employed in either sector, but not while unemployed. The basic premise is that H-skilled workers are on average more productive than L-skilled workers. As with the employer learning process, the human capital accumulation process could be modeled as a continuous learning process such that workers gradually gain skills through learning-by-doing (as modeled in Burdett et al., 2011).¹³ This process could then be calibrated so that a L-skilled worker is expected to accumulate skills and become H-skilled within K periods so that the average rate of human capital accumulation is $1/K$ per period. However, following similar arguments as with the employer learning process, we simplify the human capital accumulation process and define $\kappa = 1/K$ as the average rate of human capital accumulation and take κ as the steady-state probability that a L-skilled worker accumulates skills and becomes H-skilled. Therefore, we assume that human capital accumulation is a stochastic process such that at every moment a L-skilled worker accumulates human capital and becomes H-skilled with probability κ_F if employed in the formal sector and with probability κ_I if employed in the informal sector.

To illustrate the importance of skill accumulation in our model environment, consider the value function for an informal-firm matched with a worker with known skill level $k \in \{H, L\}$

¹³Suppose that p is the revealed worker productivity and that a worker's productivity increases at rate $\varrho > 0$ due to learning-by-doing. Then after d periods of work experience, the worker's productivity will be $p' = pe^{\varrho d}$. Then, given the distribution of skill levels between the lower bound p_L and the upper bound p_H we could find the average number of periods K that it takes a worker to reach skill level p_H .

and with match quality x :

$$(5) \quad rJ_{Ik}(x) = p_{Ik}x - w_{Ik}(x) + (\lambda_I + \eta f_F \bar{G}(C_{Fk}) + \tau)[V_I - J_{Ik}(x)] \\ + \kappa_I[J_{IH}(x) - J_{Ik}(x)] + \pi[V_I - T - J_{Ik}(x)]$$

where p_{Ik} , $k \in \{H, L\}$, denotes labor productivity for H-skilled and L-skilled workers, respectively, when employed in the informal sector. The value function indicates four reasons for termination of informal-sector jobs identical to those for informal-sector firms matched with newcomers, that is: (i) exogenous job destruction, (ii) the worker finds a formal-sector job and quits, (iii) the worker permanently exits the labor market, and (iv) the job is detected by the authorities and terminated. The losses experienced by firms after these job terminations are analog to those discussed above: $[V_I(x) - J_{Ik}(x)]$ for the first three cases and $[V_I(x) - T - J_{Ik}(x)]$ for the fourth case, for $k \in \{N, H, L\}$. Notice that when $k = L$ the value function indicates that L-skilled workers have the chance to accumulate skills and become H-skilled with probability κ_I . If this is the case the firm has a gain of $[J_{IH}(x) - J_{IL}(x)]$, which is positive if $p_{IH} > p_{IL}$ and the skill accumulation process would not lead to job destruction.

Similarly, the value function for a formal-sector firm matched with a worker with known skill level $k \in \{H, L\}$ and with match quality x is given by:

$$(6) \quad rJ_{Fk}(x) = p_{Fk}x - w_{Fk}(x) + \lambda_F[V_F - D - J_{Fk}(x)] + \kappa_F[J_{FH}(x) - J_{Fk}(x)] + \tau[V_F - J_{Fk}(x)]$$

where p_{Fk} , $k \in \{H, L\}$, denotes labor productivity for H-skilled and L-skilled workers, respectively, when employed in the formal sector. The value function indicates two reasons for job destruction for workers with a known skill level: (i) jobs can be exogenously destroyed and (ii) the worker permanently exits the labor market. In the first case the firm observes a loss of $[V_F - D - J_{Fk}(x)]$, which involves incurring the firing cost D , and in the second case a loss of $[V_F - J_{Fk}(x)]$, for $k \in \{H, L\}$. The value function for L-skilled workers, i.e. $k = L$, indicates that with probability κ_F the L-skilled worker can become H-skilled and in this case the firm experiences a positive gain of $[J_{FH}(x) - J_{FL}(x)]$ assuming that $p_{FH} > p_{FL}$.

The last part of the firm's side is the value of creating a vacancy, which is given by:

$$(7) \quad rV_j = -h_j + q_j \varphi_{jL} \int_{C_{jL}}^1 [J_{jL}(x) - V_j] dG(x) + q_j \varphi_{jH} \int_{C_{jH}}^1 [J_{jH}(x) - V_j] dG(x) \\ + q_j(1 - \varphi_{jL} - \varphi_{jH}) \int_{C_{jN}}^1 [J_{jN}(x) - V_j] dG(x)$$

where h_j is the recruitment cost in sector $j \in \{F, I\}$, q_j is the contact probabilities defined

above, and only contacts with match quality $x > C_{jk}$ result in a match being created, for $j \in \{F, I\}$ and $k \in \{N, H, L\}$. Equation (7) shows how firms do not open skill-specific vacancies for H-skilled, L-skilled, or newcomers. That is, firms open a vacancy and adjust their hiring standard depending on the skill level of the worker they meet. This equation also shows how the firm's value of a vacancy in sector j depends on the fraction of job seekers of a given skill level, φ_{jk} , for sectors $j \in \{F, I\}$ and skill levels, $k \in \{N, H, L\}$, these values are determined by the steady-state worker flows discussed in Appendix C.

2.5 Worker's Value Functions and Wage Determination

The supply side of the labor market considers the utility that workers derive from each employment condition given their skill level. Define U_k and $W_{jk}(x)$ as the present discounted value of the expected income stream of, respectively, an unemployed and an employed worker of skill level k in sector j , for $j \in \{F, I\}$ and $k \in \{N, H, L\}$. The value of unemployment for a worker with skill level $k \in \{N, H, L\}$ is given by:

$$(8) \quad rU_k = z_k + f_I \int_{C_{Ik}}^1 [W_{Ik}(x) - U_k] dG(x) + f_F \int_{C_{Fk}}^1 [W_{Fk}(x) - U_k] dG(x) - \tau U_k$$

where z_k is the flow utility from unemployment, and (f_F, f_I) are the contact probabilities defined above. Just like with the firm's side, a worker will only take a job if the match quality with the firm is high enough given the worker's skill level and the firm's employment sector.

Recall that all newcomers enter the labor market through unemployment and once they find a job in sector $j \in \{F, I\}$ with match quality x the values of being employed are given by:

$$(9) \quad \begin{aligned} rW_{IN}(x) = & w_{IN}(x) + (\lambda_I + \pi)[U_N - W_{IN}(x)] + \eta f_F \int_{C_{FN}}^1 [W_{FN}(x') - W_{IN}(x)] dG(x') \\ & + \sigma_I \nu \left(\Gamma_{IL}(x)[W_{IL}(x) - W_{IN}(x)] + (1 - \Gamma_{IL}(x))[U_L - W_{IN}(x)] \right) \\ & + \sigma_I (1 - \nu) \left(\Gamma_{IH}(x)[W_{IH}(x) - W_{IN}(x)] + (1 - \Gamma_{IH}(x))[U_H - W_{IN}(x)] \right) \\ & - \tau W_{IN}(x) \end{aligned}$$

$$\begin{aligned}
(10) \quad rW_{FN}(x) = & w_{FN}(x) + \lambda_F[U_N - W_{FN}(x)] \\
& + \sigma_F \nu \left(\Gamma_{FL}(x)[W_{FL}(x) - W_{FN}(x)] + (1 - \Gamma_{FL}(x))[U_L - W_{FN}(x)] \right) \\
& + \sigma_F(1 - \nu) \left(\Gamma_{FH}(x)[W_{FH}(x) - W_{FN}(x)] + (1 - \Gamma_{FL}(x))[U_H - W_{FN}(x)] \right) \\
& - \tau W_{FN}(x)
\end{aligned}$$

where $w_{IN}(x)$ and $w_{FN}(x)$ denote wages paid by informal and formal jobs, respectively. With probability λ_j , $j \in \{F, I\}$, the job is exogenously destroyed in which case newcomers become unemployed. Additionally, an informal-sector job can be exogenously destroyed if the firm is caught by the authorities which happens with probability π . Informal-sector workers keep searching for jobs in the formal sector and with probability ηf_F they contact a formal-sector firm with a vacancy; at this point the firm and the informal-sector worker draw a new match quality x' from distribution $G(\cdot)$ and if the match quality x' is larger than the cutoff for newcomers in the formal sector, C_{FN} , the newcomer quits the informal-sector job and moves into the formal-sector job.¹⁴ When workers move from the informal to the formal sector they have a gain of $[W_{FN}(x') - W_{IN}(x)]$, given that $x' > C_{FN}$.

The skill level of a newcomer can be revealed in the formal or informal sectors with probability σ_j , for $j \in \{F, I\}$, and depending on the current match quality the worker and the firm renegotiate their wage contract or separate. This is reflected in the value function with the indicator function $\Gamma_{jk}(x) = \mathbf{1}\{x \geq C_{jk}\}$ for $j \in \{F, I\}$ and $k \in \{H, L\}$. If the current match quality is higher than the corresponding cutoff given the worker's skill level then the contract is renegotiated and the worker has a gain of $[W_{jk}(x) - W_{jN}(x)]$, $j \in \{F, I\}$ and $k \in \{H, L\}$. If the current match quality is lower than the corresponding cutoff the match is destroyed and the worker has loss of $[U_k - W_{jN}(x)]$, $j \in \{F, I\}$ and $k \in \{H, L\}$. Finally, if a worker permanently exits the labor market before their skill level is revealed that worker suffers a loss of $W_{jN}(x)$, $j \in \{F, I\}$.

Once a worker's skill level is revealed, workers found to be L-skilled can accumulate human capital while employed in the formal or the informal sector. The value of employment in the

¹⁴Notice that both unemployed and informal newcomers have the same hiring standard, C_{FN} , regardless of their employment status. We follow Pissarides (1994) by assuming that once a match between a worker and a firm occurs the threat point of the worker is unemployment. This is based on the assumption that wage contracts are negotiated continuously. Dolado et al. (2009) and Bosch and Esteban-Pretel (2012) adopt this assumption in their matching models. See page 206 of Dolado et al. (2009) and footnote 20 of Bosch and Esteban-Pretel (2012).

formal or the informal sector for a worker of skill level $k \in \{H, L\}$ is given by:

$$(11) \quad rW_{Ik}(x) = w_{Ik}(x) + (\lambda_I + \pi)[U_k - W_{Ik}(x)] + \eta f_F \int_{C_{Fk}}^1 [W_{Fk}(x') - W_{Ik}(x)] dG(x') \\ + \kappa_I [W_{IH}(x) - W_{Ik}(x)] - \tau W_{Ik}(x)$$

$$(12) \quad rW_{Fk}(x) = w_{Fk}(x) + \lambda_F [U_k - W_{Fk}(x)] + \kappa_F [W_{FH}(x) - W_{Fk}(x)] - \tau W_{Fk}(x)$$

where the elements of the value functions are very similar to those of newcomers. The difference with the value functions for newcomers is that now that workers' skills have been revealed, L-skilled workers accumulate human capital with probability κ_F and κ_I when employed in the formal and informal sectors, respectively. Workers who accumulate human capital have a gain of $[W_{jH}(x) - W_{jL}(x)]$ for $j \in \{F, I\}$. Notice that the gain associated with human capital accumulation only enters the value function of L-skilled workers.

The final component of the model is given by wages. Our model has a similar structure as Bosch and Esteban-Pretel (2012) and Cano-Urbina (2015) and we use a similar strategy for wage determination as in those models. In particular, wages for all workers in the labor market are determined according to a surplus-sharing rule that entitles workers to a fraction β of the match surplus. The match surplus in sector $j \in \{F, I\}$ for skill level $k \in \{N, H, L\}$ is given by $S_{jk}(x) = W_{jk}(x) - U_k + J_{jk}(x) - V_j$, and so the surplus-sharing rule dictates that $[W_{jk}(x) - U_k] = \beta S_{jk}(x)$. The wage equations for each combination of worker skill level and employment sector are described in Appendix A.

3 Equilibrium

The steady state equilibrium for our model consists of three main blocks of equations. The first block considers the determination of hiring standards in both the formal and informal sectors. The second block determines the sector-specific vacancy creation conditions. The third block determines the equilibrium flow of workers between each of the labor market states presented in Appendix C.

3.1 Endogenous Hiring Standards

When workers and firms meet they create a working relationship if and only if they draw a match quality that is higher than a given cutoff. To solve for these endogenous hiring cutoffs, we evaluate the firms' value functions (equations (3) - (6) above) at the specific

match quality which brings their value to zero. The three hiring standards for the informal sector are given by:

$$(13) \quad p_{IH}C_{IH} = w_{IH}(C_{IH}) + \pi T$$

$$(14) \quad p_{IL}C_{IL} = w_{IL}(C_{IL}) + \pi T - \kappa_I J_{IH}(C_{IL})$$

$$(15) \quad p_{IN}C_{IN} = w_{IN}(C_{IN}) + \pi T - \sigma_I \nu \Gamma_{IL}(C_{IN}) J_{IL}(C_{IN}) - \sigma_I (1 - \nu) \Gamma_{IH}(C_{IN}) J_{IH}(C_{IN})$$

where $w_{Ik}(C_{Ik})$ for $k \in \{N, H, L\}$ represent the reservation wage in the informal sector for newcomers, H-skilled, and L-skilled workers because those are the wages corresponding to a match quality of $x = C_{Ik}$.¹⁵ The right-hand side of equation (13) shows that the hiring standard for H-skilled informal workers depends on the reservation wage of H-skilled in the informal sector and the likelihood the firm is detected and punished for operating informally.¹⁶ Intuitively, increases in either the reservation wage or the likelihood and size of punishment results in firms becoming more selective in hiring, leading to an increase in C_{IH} . Similar mechanisms are also present in equations (14) and (15). However, these expressions must also account for human capital accumulation and employer learning. Specifically, the last term in equation (14) represents the gain to firms when workers transition from L-skilled to H-skilled. The last two terms in equation (15) represent the gain to firms when a newcomer's type is revealed. Therefore, increases in the rate of human capital accumulation and employer learning can be seen to relax hiring standards in the informal sector.

The three hiring standards for the formal sector are given by:

$$(16) \quad p_{FH}C_{FH} = w_{FH}(C_{FH}) + \lambda_F D$$

$$(17) \quad p_{FL}C_{FL} = w_{FL}(C_{FL}) + \lambda_F D - \kappa_F J_{FH}(C_{FL})$$

$$(18) \quad p_{FN}C_{FN} = w_{FN}(C_{FN}) + [\lambda_F + \sigma_F \nu (1 - \Gamma_{FL}(C_{FN})) + \sigma_F (1 - \nu) (1 - \Gamma_{FH}(C_{FN}))] D \\ - \sigma_F \nu \Gamma_{FL}(C_{FN}) J_{FL}(C_{FN}) - \sigma_F (1 - \nu) \Gamma_{FH}(C_{FN}) J_{FH}(C_{FN})$$

The right-hand side of equation (16) shows that the hiring cutoff for H-skilled formal workers depends on the reservation wage of H-skilled in the formal sector and the expected firing costs the firm must pay. Increases in either the reservation wage or expected firing costs will result in tighter hiring standards and an increase in C_{FH} . Similar mechanisms operate for L-skilled workers and newcomers, but for these workers we must consider the impact of human capital accumulation and employer screening. The last term in equation (17) captures the benefit

¹⁵The reservation wages follow from the steady state wage equations presented in Appendix A, but evaluated at $x = C_{Ik}$ for $k \in \{N, H, L\}$.

¹⁶To ease exposition, the equations for hiring standards presented in the main text of the paper, equations (13) - (18), are not reduced. See Appendix B for reduced (closed-form) versions of the hiring standards.

firms receive when L-skilled workers gain human capital and become H-skilled. Therefore, increases in the rate of human capital accumulation in the formal sector, κ_F , will relax hiring standards and lead to a reduction in C_{FL} . Equation (18) shows that type revelation may either tighten or relax hiring standards depending on the values of $\Gamma_{FL}(C_{FN})$ and $\Gamma_{FH}(C_{FN})$ and on the distribution of workers across types, ν . Specifically, if having a worker's type revealed is likely to lead to a job separation, then this will tighten hiring standards among newcomers in the formal sector and lead to an increase in C_{FN} . If workers are very likely to maintain their employment relationship after having their type revealed, then increases in this rate are likely to relax hiring standards and reduce C_{FN} .

3.2 Job Creation Conditions

The next two equilibrium conditions are determined by the job creation conditions. In equilibrium, free-entry in both the formal and informal sectors imply that all gains from an additional vacancy are exploited, so that $V_F = V_I = 0$. Then from equation (7) the job creation condition for firms in sector $j \in \{F, I\}$ is given by:

$$(19) \quad \frac{h_j}{q_j} = \varphi_{jL} \int_{C_{jL}}^1 J_{jL}(x) dG(x) + \varphi_{jH} \int_{C_{jH}}^1 J_{jH}(x) dG(x) + (1 - \varphi_{jL} - \varphi_{jH}) \int_{C_{jN}}^1 J_{jN}(x) dG(x)$$

which indicates that the expected recruitment cost in both sectors should equal the expected profit of a match given the distribution of workers in the labor market and the distribution of match quality $G(\cdot)$.

3.3 Core Steady-State System

The equilibrium of the model is represented by three blocks of equations: (i) Hiring standards: equations (13)-(18), (ii) Job creation conditions: equation (19) for $j \in \{F, I\}$, and (iii) Steady state worker flows: equations (1c)-(9c) in Appendix C. However, the core steady-state system for our model consists of equations (13)-(19), where (19) summarizes two conditions, and depends on the following endogenous variables: C_{FH} , C_{FL} , C_{FN} , C_{IH} , C_{IL} , C_{IN} , θ_F , and θ_I . The equilibrium worker flows and matching probabilities can be determined recursively given values for the sector-specific hiring standards and labor-market tightnesses. Given this core system, we calibrate our model to replicate key features of the Mexican labor market during the sample period 2005-2018, with a focus on young less educated workers. In the following sections we provide a detailed overview of the survey data used to determine our empirical targets and the methodology that we use to pin down key parameter values.

4 Data: The ENOE

To bring our model to the data, we turn to the Occupation and Employment Survey (ENOE: *Encuesta Nacional de Ocupación y Empleo*), a household survey from Mexico with detailed information on income and employment status. The ENOE is a rotating panel where households are visited five times over the course of a year.¹⁷ As such, the ENOE provides information quarterly, with 20% of the sample being rotated out and replaced by new households each quarter. During each visit, information is recorded regarding the demographics (e.g., education, age, marital status) and main and secondary jobs (e.g., hours, earnings, benefits, position, firm-size, industry, occupation and job tenure) of each family member over the age of 12. While the information provided in the ENOE is at the individual level, it is reported by a single member of the household and the reporting member may change across subsequent visits.¹⁸ For more information on the ENOE, see INEGI (2005, 2007).

To focus on young, less-educated individuals, we restrict the sample to include only those age 16 to 30 with less than 12 years of education and who are not currently enrolled in school. The education restriction is uncontroversial and is similar to focusing on those with less than a high school education within the United States. However, one could argue that removing individuals less than 16 years of age from the sample ignores a sizable fraction of informal workers, as many individuals below the age of 16 work in the informal sector. The issue here is that 16 is the minimum age to work legally in Mexico (see Congress, 1970) and our primary interest is on the transition between informal and formal employment. Therefore, individuals younger than 16 are not eligible to transition to the formal sector, even if a position was offered, and as such, we remove them from our sample. The upper limit of the age range is set to 30 as this seems to be the age at which transitions from the informal to the formal sector have reached a plateau (see Figure 2).

Along with the age and education restrictions described above, we also restrict our sample to only include males. This gender restriction was put in place as women may choose to participate in the informal sector for a variety of different reasons, e.g., job flexibility, work-

¹⁷For example, a household visited for the first time in January 2005 will be visited a second time in April of 2005, a third time in July of 2005, a fourth time in October 2005, and a fifth time in January of 2006. Not all households are visited for the first time in January, they could be visited for the first time in any other month of the year.

¹⁸We explored the possibility of using the MOTRAL, which is another survey from Mexico where we can follow the duration of the individual in multiple jobs over a period of five years. While the MOTRAL also collects data on salary and whether individuals receive health benefits like IMSS or ISSSTE, only one salary and one IMSS affiliation is reported for the entire period of the job, regardless of the job duration. That is, the survey reports jobs with a duration of two or three years but only one wage measure for that job. As such, we are unable to learn anything about the duration of time between wage changes within a given job, which is crucial for our identification strategy.

life balance, child rearing (Arias and Maloney, 2007) and as such, formal employment may not be a strictly preferable outcome. And finally, we use data from the first quarter of 2005 to the fourth quarter of 2018 and we restrict the sample to urban areas.¹⁹

We classify individuals as employed formally or informally using institutional details from Mexico. Specifically, we make use of Mexican laws regarding the provision of health benefits to workers in order to classify each individual across employment sectors. In Mexico, when a worker is hired the employer is required to register the worker in the IMSS or the ISSSTE, with non-complying firms incurring a penalty if caught.²⁰ These institutions provide workers with benefits including health insurance, daycare services for children, life insurance, disability pensions, etc. (Levy, 2007).²¹ Within the survey, we observe if individuals have access to the health services provided by the IMSS or the ISSSTE. We label a worker as employed formally if he is both salaried and has access to either of these health services, and as employed informally if he is salaried but does not have access to either of these health services. Notice that the self-employed are not included in our definition of the informal sector as we restrict attention to salaried employees. Summary statistics for both the formal and informal sectors can be found in Appendix D

5 Estimating Employer Learning and Human Capital Accumulation Rates

5.1 Using Model Predictions to Estimate Parameters

Parameters $(\sigma_F, \kappa_F, \sigma_I, \kappa_I)$ govern the sector-specific rates of employer screening and human capital accumulation in our model. We use duration data from the ENOE to estimate these parameters. The estimation is based on our model’s predictions regarding wage changes. Once calibrated, our model predicts that wage changes will occur for one of two reasons:

- P1. A newcomer’s skill level is revealed and the wage is adjusted up or down.
- P2. A L-skilled worker becomes H-skilled and the wage is increased.

¹⁹We include all the cities that are statistically self-represented in the ENOE. This includes 32 cities, all except one with a population greater than 100,000. The one self-represented city with a population less than 100,000 is Tlaxcala. It is classified as a city with a population between 15,000 and 99,999. According to the INEGI, the Statistical Bureau in Mexico in charge of the Census (and the ENOE), in 2010 the population of the Tlaxcala was 89,795 (see <http://www3.inegi.org.mx/sistemas/mexicocifras/default.aspx>).

²⁰IMSS is the acronym in Spanish for the Mexican Institute of Social Security and ISSSTE is the acronym in Spanish for the Institute of Security and Social Services for the State’s Workers.

²¹While both the worker and the employer must pay fees to fund these institutions, the portion paid by the employer is much higher than that paid by the worker.

With these predictions in mind, we determine whether wage changes in the data can be attributed to skill revelation or skill accumulation by restricting our attention to two specific subsamples based on the distribution of hourly wages in the data on first ENOE interview.

S1. Workers who are on the middle of wage distribution on first interview

S2. Workers who are on the bottom of wage distribution on first interview

For the case of employer learning, the calibrated model predicts that, in both sectors, the wage of a newcomer is between that of L-skilled and H-skilled, that is, $w_{jL}(x) < w_{jN}(x) < w_{jH}(x)$ for $j \in \{F, I\}$. Then, we assume that workers who experience a wage change attributed to skill revelation must start in the middle of the wage distribution, as they are the newcomers, and then experience a wage change that moves them to the top or the bottom of the wage distribution if they are revealed as H-skilled or L-skilled, respectively. These movements are described in panel (a) of Figure 3.

For the case of human capital accumulation, the calibrated model predicts that, in both sectors, the wage of a L-skilled worker increases when the worker accumulates skills and becomes H-skilled, and so $w_{jL}(x) < w_{jH}(x)$ for $j \in \{F, I\}$. Then, we assume that workers who experience a wage change attributed to skill accumulation must start in the bottom of the wage distribution, as they are already revealed as L-skilled, and then experience a wage change that moves them to the top of the distribution once they accumulate skills and become H-skilled. This movement is described in panel (b) of Figure 3.

Therefore we identify wage changes attributed to:

- (i) Skill revelation using model's prediction P1 and subsample S1 (as in Figure 3(a)).
- (ii) Skill accumulation using model's prediction P2 and subsample S2 (as in Figure 3(b)).

Once we have identified these wage changes, we measure the time from the first interview to the interview in which an individual experiences a wage change that can be attributed to screening or the accumulation of skills. We refer to this length of time as the *duration of wage changes*. Appendix E.1 provides more details about these duration measures. With these duration data we estimate hazard functions to obtain estimates of σ_F , σ_I , κ_F , and κ_I . First, we estimate two hazard functions using the durations of wage changes for workers that start in the middle of the wage distribution, to obtain estimates of σ_F and σ_I . Next, we estimate two hazard functions using the durations of wage changes for workers who start in the bottom of the wage distribution, to obtain estimates of κ_F and κ_I .

Figure 4 presents the distribution of workers in the formal and informal sectors by age in the subsamples used to estimate σ_F , σ_I , κ_F , and κ_I . That is, the graphs presented in

Figure 3: Wage Changes used to Estimate σ and κ

(a) Attributed to Screening

(b) Attributed to Human Capital

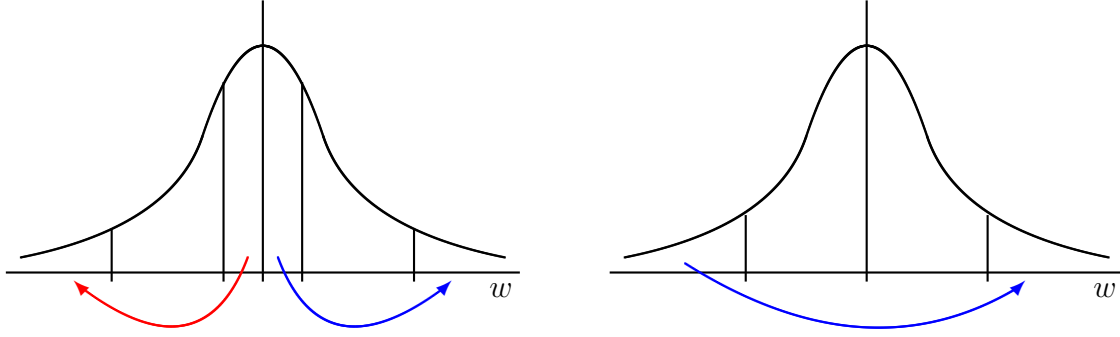


Figure 4: Distribution of Workers by Age in Subsamples to Estimate Employer Learning and Human Capital Accumulation Rates.

(a) Subsample to Estimate σ_F and σ_I

(b) Subsample to Estimate κ_F and κ_I

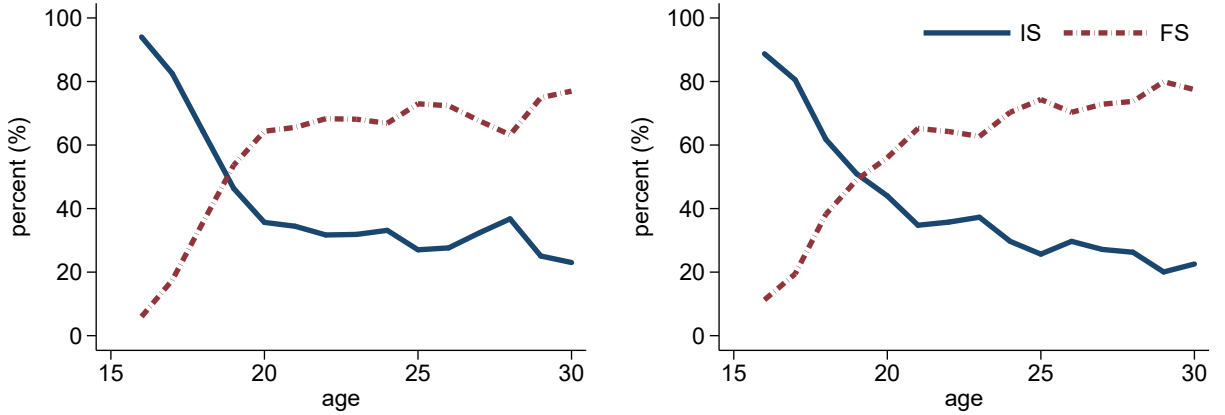


Figure 4 are intended to explore whether the general patterns observed in Figure 1, and that largely motivate this paper, are still present in the subsamples. As can be seen by the two graphs in Figure 4, the subsamples show the same patterns observed for the overall sample. Notice that the graphs for Figure 4 do not include the self-employed and the employers as Figure 1 does because we exclude these workers from our sample. Also, notice that Figure 2 cannot be replicated because the subsamples exclude workers who make transitions between the formal and informal sectors as indicated above.

5.2 Estimation Using Hazard Functions

The way we model employer learning in Section 2 assumes that in every period the skill level of a newcomer can be revealed with probability σ_j while employed in sector $j \in \{F, I\}$. Similarly, the way we model human capital accumulation assumes that in every period a L-skilled worker can accumulate skills and become H-skilled with probability κ_j while employed in sector $j \in \{F, I\}$. Therefore, both the revelation of a worker’s skill level and the accumulation of skills are Poisson processes in the theoretical framework we developed in Section 2. As a result, the durations of wage changes due to screening or human capital accumulation described in the previous section are exponentially distributed (see section 2 of chapter 5 in Lancaster, 1990). The exponential duration distribution has a constant hazard rate that does not vary with the duration spent in a given state, which is the memoryless property of the exponential distribution (Cameron and Trivedi, 2005), and that is appropriate for the way we model both skill revelation and accumulation because, in our model, skills can be revealed or accumulated with the same probability regardless of the time the individual has been employed. In Appendix E.3, we verify that the exponential function is indeed an appropriate representation of the simplified processes of employer learning and human capital accumulation from the model in the data for young less-educated workers in Mexico.

The duration of time observed before realizing a wage change that is the result of employer screening or human capital accumulation is represented by the random variable t . Appendix E.1 describes in detail how the duration variable t is constructed for each sample used to estimate σ_F , σ_I , κ_F , and κ_I . If this duration variable follows an exponential distribution, then the hazard function is a constant function of the form $\phi(t) = \phi$. When we estimate the hazard function with the sample of workers who start in the middle of the distribution in sector j , the constant hazard function is $\phi(t) = \sigma_j$, for $j \in \{F, I\}$. Similarly, when we estimate the hazard function with the sample of workers who start in the bottom of the distribution in sector j , the constant hazard function is $\phi(t) = \kappa_j$, for $j \in \{F, I\}$. As described in Appendix E.1, all samples use standardized wage measures that have been purged of observable individual characteristics, hence we do not include covariates in the estimation of these hazard functions.

Estimates of the hazard functions are obtained by maximum likelihood estimation (MLE). In Appendix E.2, we provide all the details for the estimation of these hazard functions. The estimation results are presented in Table 1. In this table, the number of failures represent the number of individuals who made a transition to the first or last deciles, for skill revelation, or to the top two deciles, for skill accumulation. The rest of the individuals are right-censored.

Table 1 indicates that the estimates for the average rate of employer learning are $\sigma_F =$

0.0424 and $\sigma_I = 0.0544$ which are significant at the 1% level. These estimates suggest that the average amount of time it takes an employer to gather enough information about a young less-educated worker that is new to the labor market to determine if the worker is L-skilled or H-skilled is about 24 months in the formal sector and 18 months in the informal sector. Similarly the estimates for the average rate of human capital accumulation are $\kappa_F = 0.0278$ and $\kappa_I = 0.0358$ which are also significant at the 1% level. These estimates suggest that the average amount of time it takes a young less-educated worker who is L-skilled to accumulate enough human capital to become H-skilled is about 36 months if employed in the formal sector and 28 months if employed in the informal sector. Finally, since the rates reported in Table 1 are at a monthly frequency, we must scale the estimates by three before using them in our quarterly model.

Table 1: Maximum Likelihood Estimates of Employer Learning and Human Capital Accumulation Rates

	σ_F	σ_I	κ_F	κ_I
Estimate	0.0424	0.0544	0.0278	0.0358
	(0.0021)	(0.0032)	(0.0016)	(0.0025)
No. of observations	1,042	609	1,057	595
No. of failures	414	290	300	206

NOTES: Standard errors in parentheses. “No. of failures” correspond to the number of observations with duration that is not right censored. Duration data derived from the ENOE. More details on the data and the methodology used to estimate these parameters are provided in Appendix E.

6 Model Calibration

After estimating the values for σ_j and κ_j for $j \in \{I, F\}$ using the method described above, we must still determine the values for the remaining parameters of our model. To simplify this process, we assume that the parameters governing the matching functions are symmetric across employment sectors, while workers’ utility from unemployment is symmetric across skill-type (e.g., $\gamma_1^F = \gamma_1^I$; $\gamma_2^F = \gamma_2^I$; and $z = z_H = z_L$). Given that the data from the ENOE is quarterly, we calibrate our model to a quarterly frequency and set the interest rate, r , to 0.01. This restriction implies an annualized rate of return of approximately 4%, which is consistent with existing estimates for Mexico. We also set the exit rate, τ , to 0.0179 to target an average duration of approximately 14 years. This ensures that the time horizon for the agents in our model is consistent with that of subjects in our sample (e.g., individuals

age 16 to 30).²²

Following Bosch and Esteban-Pretel (2012), we set $\eta = 0.4$, implying that on-the-job search while employed in the informal sector is less efficient than traditional search from unemployment. Following the same authors, we set the parameter governing informal detection, π , to 0.005. While detection by authorities results in the termination of an informal position, we set the additional penalty firms face if caught operating informally, T , to 0 in our baseline calibration. This is consistent with a monitoring authority that is not currently undertaking actions to monitor and punish firms who hire workers informally. For our surplus sharing rule, we follow the literature and assume equal weights for firms and workers, $\beta = 0.5$. We also set the elasticity of matches with respect to unemployment γ_2 equal to 0.5.²³

In order to calibrate the remaining parameters, we must turn our attention back to our sample from the ENOE. We set the exogenous separation rates in our model, λ_I and λ_F , so that the endogenous transition rates generated by our model match those observed in the data for their respective sector (see Appendix D for ENOE transition probabilities). The coefficient in the matching function, γ_1 , is set to target an unemployment rate of approximately 9 percent, consistent with our sample from the ENOE. Next, we determine the fraction of newcomers who are L-skilled, ν , using the same data on wage changes that was used to estimate σ_F and σ_I . The parameter ν is obtained with a back-of-the-envelope calculation that uses the fraction of transitions from the middle to the top of the distribution and is explained in detail in Appendix E.4. Agent's flow utility from unemployment, z , is set to 40 percent of the average wage in the formal sector, thereby targeting the replacement rate. While Mexico does not have a comprehensive unemployment insurance program, our value of z should be interpreted as the total value of income received while unemployed, which includes, among other things, the value of home production. Vacancy posting cost in the formal sector, h_F , are set 1, while vacancy posting costs in the informal sector are set to a fraction of this value in order to target an informal employment share of approximately 48 percent. And finally, firing costs in the formal sector, D , are set so that they are approximately 10 times the average wage in the formal sector.²⁴ The calibration details are summarized in Table 2

²²While τ is often interpreted in these types of models as the probability of death, we are interpreting τ as the probability an agent transitions out of this limited segment of the labor market.

²³Our primary results are not sensitive to small variations in these parameter values.

²⁴Empirical studies examining labor costs in Mexico reveal that firing costs are quite large. Montes Rojas and Santamaría (2007) estimate the direct payments of firing costs to be in excess of 6 times quarterly wages. However, these direct costs represent a fraction of the full costs faced by firms due to labor disputes related to employee terminations. In addition to the direct costs, firms must often cover workers' court costs and pay back-wages while termination disputes are ongoing. As such, firing costs that are 10 times the average quarterly wage is within the plausible range for the full cost associated with formal separations in Mexico.

Table 2: Calibration Summary

Parameter	Symbol	Value	Target or Reference
Exog. Separation Rate FS ^a	λ_F	0.0415	ENOE Transition Rate
Exog. Separation Rate IS ^a	λ_I	0.0705	ENOE Transition Rate
Interest Rate	r	0.01	Annualized Rate of Return
Permanent Exit Rate	τ	0.0179	Worker's life = 14 years
On-Job-Search Efficiency	η	0.4	Bosch and Esteban-Pretel (2012)
Fraction L-skilled	ν	0.5203	Determined using wage changes
Matching Function Coefficient	γ_1	0.8650	Unemployment Rate
Probability of Detection	π	0.005	Bosch and Esteban-Pretel (2012)
Firing Costs	D	14.9	Montes Rojas and Santamaría (2007)
Disutility of Work	z	0.589	Replacement Rate
Formal Vacancy Cost	h_F	1.00	Normalized to 1
Informal Vacancy Cost	h_I	0.070	Target size of informal sector

^a The transition probability between formal and informal sector employment and unemployment in our model depends on the exogenous separation rate in the respective sector (λ_F or λ_I), as well as the proportion of newcomers who separate once their type is revealed as either low or high. We target this combined rate in our model based on the observed ENOE transition rate, but the primary determinant of this separation rate is the sector-specific value of λ_j , $j \in \{I, F\}$

The final six parameters are the skill and sector-specific productivity values (the p 's) in our model. These productivity values are determined using both ordinal and cardinal wage restrictions. Specifically, within each sector the productivity values are set to preserve the following ordering: $E(w_{jL}) < E(w_{jN}) < E(w_{jH})$, $j \in \{F, I\}$. The productivity values must also result in the following within skill-type ordering: $E(w_{Fk}) > E(w_{Ik})$, $k \in \{N, H, L\}$. Along with preserving the two orders of expected wages described above, the magnitude of the skill and sector-specific productivity values is set so that the average wage in the formal sector is approximately 17.6% higher than the average wage in the informal sector, which is consistent with the ENOE data.²⁵

7 Results

Given our calibration strategy, we compute the steady state solution of our model economy and recover a set of baseline results. These baseline results are then compared to the data to assess our model's empirical fit and derive results regarding unobservables, such as the endogenous hiring cutoff rules across both sector and skill-type, as well as skill-specific unem-

²⁵From Table D.1 we see that in the formal sector, average hourly earnings is 28.97 and average weekly hours worked is 50.67. This implies average weekly earnings of 1,467.91 in the formal sector. Applying the same calculation to the informal sector, we find that the average weekly earnings in the informal sector is 1,248.16. Therefore, average weekly earnings is 17.6% higher in the formal sector than in the informal sector.

ployment rates. These results are then used to determine if informal employment serves as a “port-of-entry” into the labor market for young less-educated workers. After assessing our baseline results, we consider a set of counterfactual experiments that present two alternative approaches to reducing the employment share of the informal sector. The first approach is to deregulate the formal sector, while the second approach increases the punishment for operating informally. The two policies are compared in terms of how they distort the equilibrium of the labor market.

7.1 Baseline Results

Our baseline model is found to match the data well. Table 3 presents the list of moments used in our calibration. The model does a very good job in targeting the unemployment rate, the size of the informal sector, and the average wage premium present in the formal sector. For an additional test of empirical fit, we conduct simulations to measure the share of workers by age employed across the formal and informal sectors. We take a unit measure of agents and start 10% in unemployment, 15% in formal employment and 75% in informal employment and then track them through time, using the steady state transition rates recovered earlier.²⁶ Figure 5(a) displays the results of this simulation exercise against the pattern observed in the data.²⁷ Inspection of this panel reveals that our baseline calibrated model does a very good job replicating the primary feature of the data with which we are interested and that largely motivated our study, namely, the rapid transition from informal to formal employment that occurs during the first few years of an individual’s working life.

We also simulate counterfactual versions of our baseline model where the strength of the employer screening and human capital accumulation mechanisms have been reduced by 60 percent, while holding all other model parameters fixed. This approach is consistent with the view that such a weakening of the mechanisms is associated with an exogenous shock²⁸. Figure 5(b) compares the counterfactual simulation where both mechanisms are weakened to the data. Reducing the strength of both mechanisms is found to greatly hinder the model’s ability to replicate the pattern observed in the data. Similarly, Figures 5(c) and 5(d) weaken each mechanism separately and show that both mechanisms are important for our model to

²⁶The initial distribution of agents across the unemployment and employment states was chosen so that the first period of our simulation is consistent with that observed in the data. Our findings are qualitatively consistent if we simply started with all agents in the unemployed state.

²⁷The data used to calculate the shares of informal and formal sectors are the same data used to produce Figure 1 but excluding self-employed and employers since these workers are not included in our theoretical framework and we are not able to simulate them.

²⁸Alternatively, one could conduct a related experiment where all model parameters are recalibrated to match the empirical targets. While we did not follow this strategy, our main calibration moments, such as the unemployment rate and the size of the informal sector, do not change substantially.

match the data closely.

Table 3: Summary of Targeted Moments

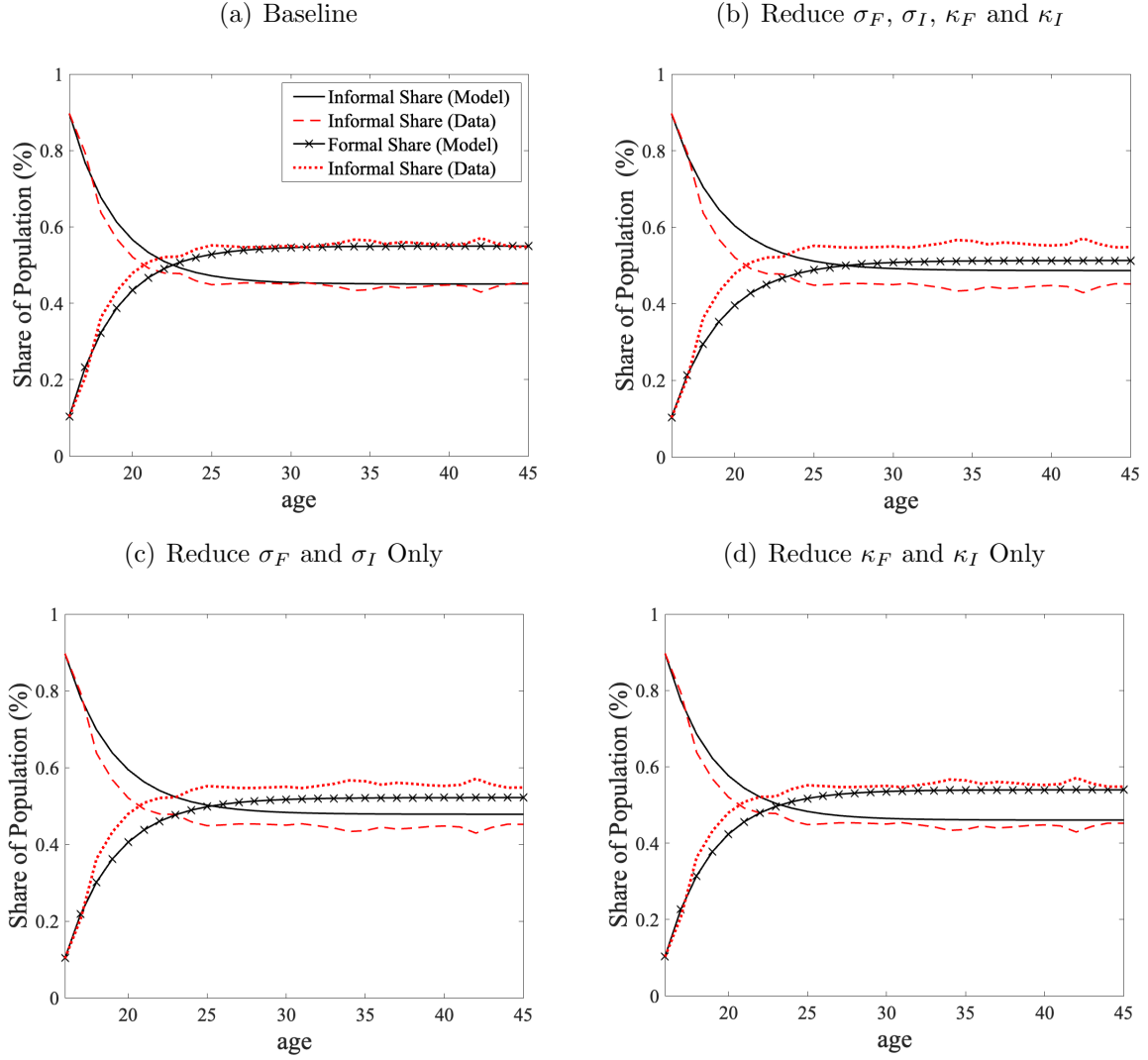
Moment	Notation	Target	Model
Unemployment rate (%)	u	9.0	9.0
Fraction of IS workers (%)	n_I	48.0	48.0
Ratio of wages	$\bar{w}_F/\bar{w}_I$	1.176	1.176
Replacement Rate	$z/\bar{w}_F$	0.40	0.40

Given the empirical fit of our baseline model, we now consider the implications of our model regarding variables of interest which are unobservable in the data, namely the equilibrium outcomes for agents across skill levels. Table 4 indicates that our baseline model makes strong predictions regarding the composition of agents across skill levels. Specifically, the first column of the table indicates that 81.33% of agents are H-skilled in the steady state equilibrium, while only 7.56% and 11.11% are L-skilled and newcomers, respectively. Given that all agents are born as newcomers and about half are initially revealed to be L-skilled, this result suggests that both employer screening and human capital accumulation have a substantial effect on the equilibrium.

To better understand the role played by the informal sector, we can look into the composition of agents across skill-type within each sector of employment, which are presented in the second and third columns of Table 4. Here we find that the vast majority of workers employed in the formal sector, 88.26%, are H-skilled. Furthermore, only 5.91% of workers employed in the formal sector are newcomers. This is in stark contrast with the skill composition found in the informal sector, where 75.08% are H-skilled and 15.80% are newcomers. The relatively high concentration of newcomers in the informal sector suggests that the informal sector in our calibrated model serves as a “port-of-entry” into the labor market for young less-educated workers. Table 5 breaks down unemployment by skill level in our baseline model. The results are intuitive as workers with known skill levels face similar unemployment rates (7.03% for H-skilled and 6.93% for L-skilled), while the unemployment rate among newcomers is substantially higher at 22.30%.

To further investigate whether the informal sector is serving as a “port-of-entry” for young less-educated workers, we turn to the equilibrium hiring standards that are derived from our baseline calibrated model. Table 6 presents these cutoffs across both sectors and for each skill-type for our baseline specification. Inspection of Table 6 indicates that the cutoffs are in fact substantively lower for the informal sector than for the formal sector. This finding is in-line with the hypothesis that formal sector employment carries with it more stringent barriers to entry which may not be present with informal sector employment. This

Figure 5: Distribution of Agents by Employment Sector (Model Vs. Data)



Notes: Figure generated by simulating the model. The data is drawn from the same ENOE sample used to generate Figure 1, only includes salaried workers, and excludes self-employed and employers.

Table 4: Distribution of Workers in Baseline Model

	Fraction of Workers		
	Labor Market	Formal Sector	Informal Sector
H-skilled	81.33	88.26	75.08
L-skilled	7.56	5.83	9.12
Newcomer	11.11	5.91	15.80

NOTES: Each column presents the fraction of workers within a particular sector or the total labor market. Therefore the numbers on each column add to one.

Table 5: Unemployment Rate by Skill Level in Baseline Model

Unemployment Rate	
H-skilled	7.03
L-skilled	6.93
Newcomer	22.30

Table 6: Hiring Standards in Baseline Model

	Formal Sector	Informal Sector
H-skilled	0.7543	0.7254
L-skilled	0.7825	0.6912
Newcomer	0.7921	0.6902

fact is most evident when comparing the hiring cutoff for newcomers across sectors (0.6902 informal sector and 0.7921 formal sector). As a consequence, newcomers in our model will more naturally flow into informal employment until they have their skill-type revealed and gain human capital, consistent with an informal sector that serves as a “port-of-entry” into the labor market.

Table 6 also shows that the ordering of cutoffs by skill level differs across sectors. Specifically, for the formal sector, it is easiest to enter as H-skill, then as L-skill, and finally as a newcomer. In contrast, for the informal sector, it is easiest to enter as a newcomer and most difficult to enter as H-skill. While the ordering for the informal sector may seem counterintuitive, it is directly related to the asset value of a worker to the informal firm. Recall that all informal workers continue searching for formal jobs. As such, H-skilled informal workers have a higher chance of transitioning to a formal firm and as such, represent a potential loss to informal firms.

7.2 Counterfactual Policy Experiments

While our baseline results cast informal sector employment opportunities in a positive light, with results suggesting that informal positions serve as a “port-of-entry” into the labor market, governments and the population may seek to reduce the size of the informal sector for a variety of other reasons.²⁹ Regardless of the motive, it is important to understand how policies that limit the size of the informal sector impinge on labor market outcomes. We consider two alternative policies which alter the size of the informal sector. The first policy adjusts the size of expected firing costs in the formal sector, thereby reducing the regulatory burden of operating formally. In contrast, the second policy experiment increases

²⁹For example, informal positions may be less productive and thus wasteful, or governments may want to capture the lost tax revenue due to informal employment contracts.

the expected penalty or fine firms must pay if they are caught operating informally, and can thus be viewed as the government “cracking down” on informal participation. By comparing the impact of these counterfactual policy experiments, we are able to determine if one policy dominates the other in terms of equilibrium outcomes within the labor market (e.g., the effect on unemployment, hiring standards, worker’s skills distributions, etc.).

To conduct our policy experiments, we re-solve the steady state of our model under a wide range of values for the expected firing costs in the formal sector and the expected fine in the informal sector. Both policies are intended to reduce the size of the informal sector, and we choose a 30% reduction in the size of the informal sector relative to baseline as the bound for the policy experiments. While the figures below describe how variables change smoothly with adjustments in the underlying policy variables, our primary point of comparison will be between the baseline steady state and the steady state consistent with a 30% reduction in the size of the informal sector. Figures 6(a) and 6(b) demonstrate how sensitive the size of the informal sector is to changes in the expected firing cost and the expected fine, respectively, while Figures 6(c) and 6(d) report the response for unemployment³⁰. Reducing the expected firing cost so that the size of the informal sector falls by 30%, is found to reduce aggregate unemployment by approximately 6%. In contrast, increasing the expected fine by a similar magnitude (e.g., to also reduce the size of the informal sector by 30%) is found to increase aggregate unemployment by approximately 15%.

While the aggregate results presented above are important, we are also interested in how such policies impact agents’ skill revelation and human capital accumulation processes, as well as unemployment across skill types. Figure 7(a) shows that reducing the expected firing cost lowers the fraction of H-skilled workers and raises the fraction of L-skilled workers and newcomers in equilibrium. A similar pattern is observed when the expected fine is increased, though the magnitude of change is now generally larger (see Figure 7(b)). Reducing (increasing) the expected firing cost (fine) to lower the size of the informal sector by 30% is found to increase the share of newcomers by approximately 2% (7%), increase the share of L-skilled workers by approximately 3% (3%), and reduce the share of H-skilled workers by approximately 0.5% (2%). Figures 7(c) and 7(d) present the impact of lowering the expected firing cost and raising the expected fine on skill-specific unemployment, respectively. While lowering the expected firing cost reduced unemployment overall, Figure 7(c)

³⁰Note that the x -axis in our policy experiment figures differs depending on the policy experiment in question (e.g., expected firing cost or expected fine; both as a share of the average formal sector wage). One policy experiment requires the expected firing cost to fall, while the other requires the expected fine to rise. Hence, results from reducing the expected firing cost should be read from right to left, while results from increasing the expected fine should be read from left to right. The same is true for Figures 7-10 describing other outcomes from the policy experiments.

shows that unemployment falls more for agents with known skill levels (H-skill or L-skill) than for newcomers. Similarly, while the aggregate results show that raising the expected fine increases unemployment overall, Figure 7(d) shows that unemployment rises more for newcomers and L-skilled workers and less for H-skilled workers. The results so far indicate that reducing the expected firing cost dominates increasing the expected fine as the former policy lowers unemployment for all agents. However, both policies are found to erode the skill revelation and human capital accumulation mechanisms present in the labor market. Next, we consider other variables within the model that can help shed light on the underlying mechanisms behind these findings.

Intuitively, the two policies reduce the size of the informal sector by simultaneously incentivizing the creation of formal jobs and disincentivizing the creation of informal jobs. However, the two policies are found to impact unemployment, both in aggregate and across skill types, in different ways, with a lower expected firing cost reducing unemployment and a higher expected fine increasing unemployment. As such, reducing the expected firing cost results in more formal-sector job gains than informal-sector job losses, while the same cannot be said for increasing the expected fine. Recall that in our model job gains or losses may result from either changes in vacancy creation or changes in hiring standards (or both). Therefore, we must consider how the policies influence these margins to better understand their impact on unemployment.

Figures 8(a) and 8(b) show that vacancy creation in the formal sector increases as both policies are intensified, with formal sector vacancies rising by approximately 50% (35%) from baseline when the expected firing cost (fine) is reduced (increased). Similarly, Figures 8(c) and 8(d) show that vacancy creation in the informal sector falls as both policies are intensified, with informal sector vacancies falling by approximately 27% (35%) from baseline when the expected firing cost (fine) is reduced (increased). Taken together, the results presented in Figure 8 show that formal sector vacancy creation rises by more, and informal sector vacancy creation falls by less, when the expected firing cost is lowered than when the expected fine is raised. This provides a partial explanation for why reducing the expected firing cost results in lower unemployment. However, as described above, we must also consider how the policies influence the endogenous hiring rules across sector and skill type to fully understand how the policies impact the flow of workers in the labor market. Figures 9(a) and 9(b) show that both policies reduce hiring standards in the formal sector, with small variation in the extent of the adjustment depending on worker skill level. In contrast, Figures 9(c) and 9(d) show that the two policies impact hiring standards in the informal sector differently. Reducing the expected firing cost is found to lower hiring standards in the informal sector, while increasing the expected fine raises hiring standards in the informal sector. The finding that

raising the expected fine generally raises hiring standards in the informal sector is intuitive as firms operating informally must become more selective in hiring as they now must factor in a significantly larger expected fine. So, while increasing the expected fine provides an incentive for creating additional formal-sector jobs, this margin is dominated by a reduction in job creation in the informal sector, thereby leading to a higher equilibrium unemployment rate. This issue is not present when the expected firing cost is reduced.

Next, we consider how the policies impact the skill-specific employment shares within each sector. Figures 10(a) and 10(b) show that both policies result in a small reduction in the employment share of H-skilled workers and significantly increase the employment share of L-skilled workers and newcomers. Specifically, reducing (increasing) the expected firing cost (fine) reduces the employment share of H-skilled workers by about 2.5% (2.5%) and increased the employment share of L-skilled workers and newcomers by about 15% (15%) and 22.5% (25%), respectively. We find that the policies have a similar impact on the employment shares in the informal sector (see Figures 10(c) and 10(d)). Taken at face value, these results suggest that both policies improve the career prospects of newcomers and L-skilled workers by increasing their employment share in both sectors. However, one must keep in mind that our previous results showed that the policy change also impacted the equilibrium skill distribution, resulting in more newcomers and L-skilled workers relative to baseline. Therefore, the increase in employment share of newcomers and L-skilled workers is driven in part by the greater prevalence of both types of agents in the post-policy equilibrium.

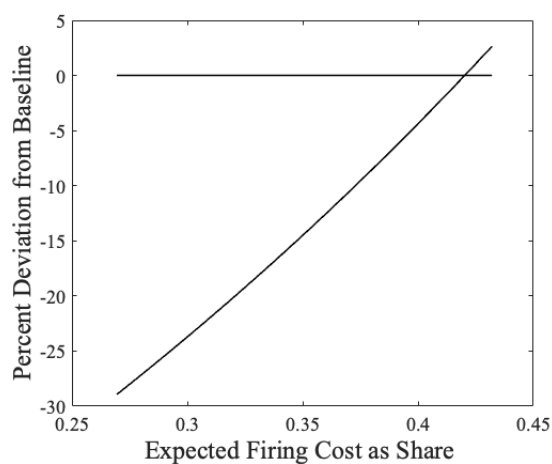
The results of the counterfactual policy experiments can be summarized as follows. While both lowering the expected firing cost and raising the expected fine incentivize the creation of formal jobs and disincentivize the creation of informal jobs, only reducing the expected firing cost generates a positive employment effect. As such, reducing firing costs can be viewed as a more favorable policy if one is concerned with reducing the size of the informal sector. However, we find that this policy still results in a larger proportion of newcomers and L-skilled workers in equilibrium. This finding suggests that policies that reduce the size of the informal sector may also unintentionally reduce the rates of transition between skill-types for young less-educated workers. Additional policies to bolster employer screening and human capital accumulation within the formal sector for young, less-educated workers should also be considered.

8 Conclusion

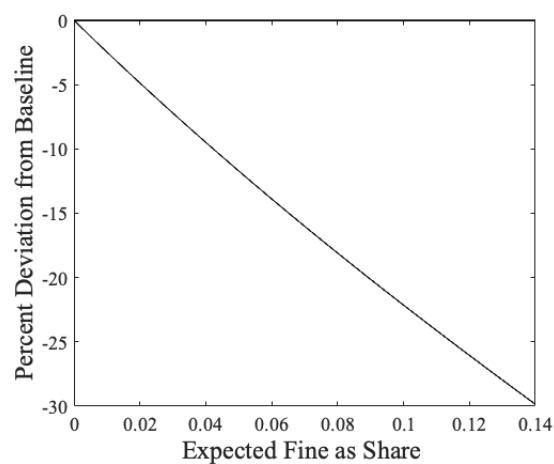
The empirical evidence suggests that informal labor participation may play an important and transient role for young less-educated workers in Mexico. To analyze this issue, we

Figure 6: Effects of Policy Experiments on Size of Informal Sector and Unemployment

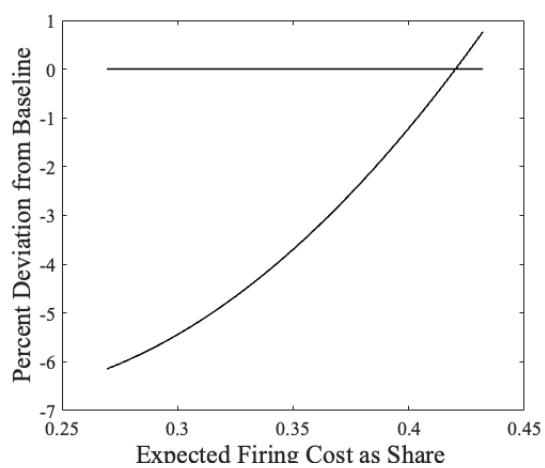
(a) Firing Costs: Size of IS



(b) Informality Penalty: Size of IS



(c) Firing Cost: Unemployment



(d) Informality Penalty: Unemployment

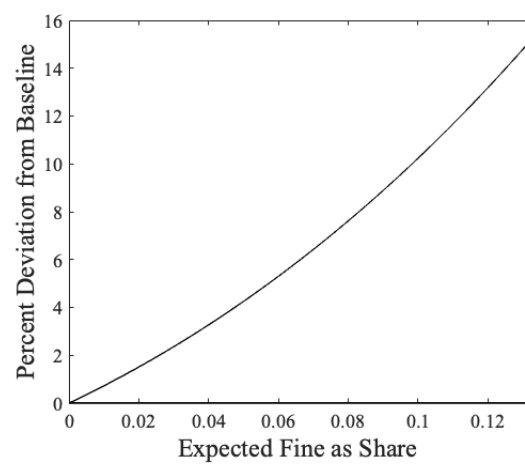
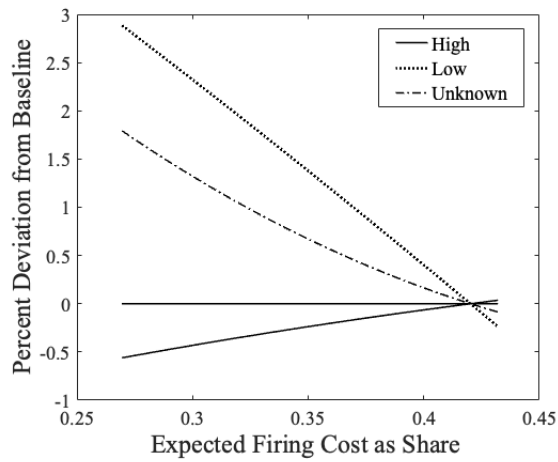
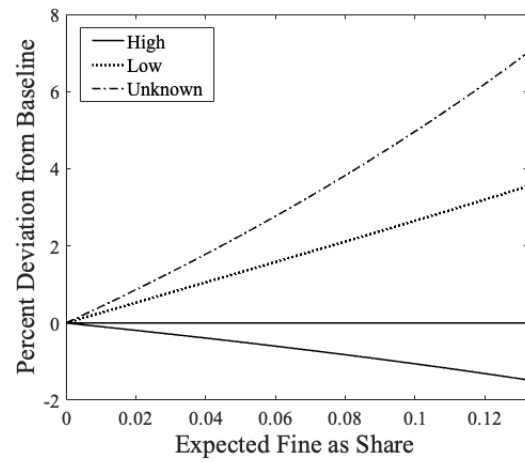


Figure 7: Effects of Policy Experiments on Aggregate Skill Composition and Unemployment by Skill Level

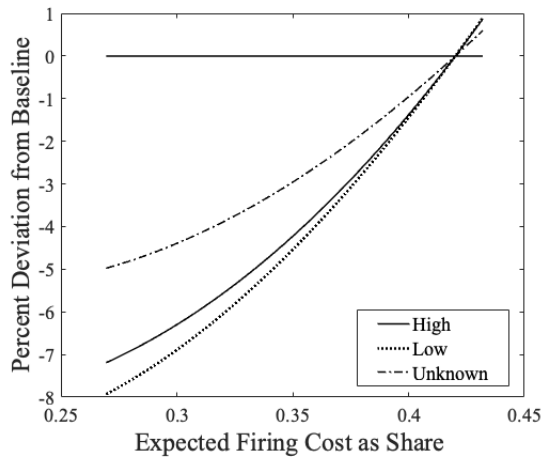
(a) Firing Costs: Skill Composition



(b) Informality Penalty: Skill Composition



(c) Firing Costs: Unemployment by Skill



(d) Informality Penalty: Unemployment by Skill

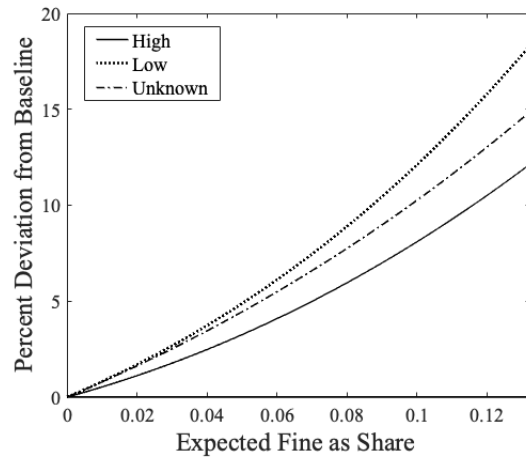
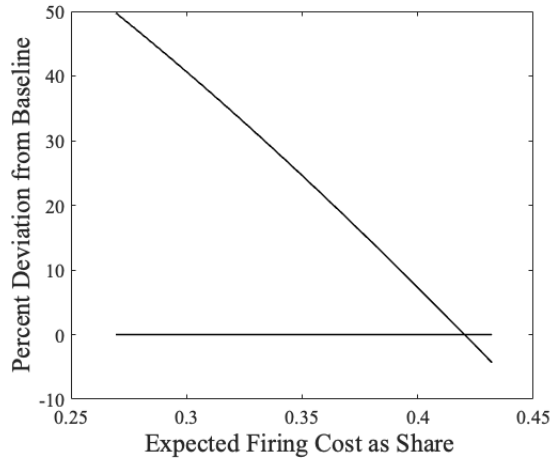
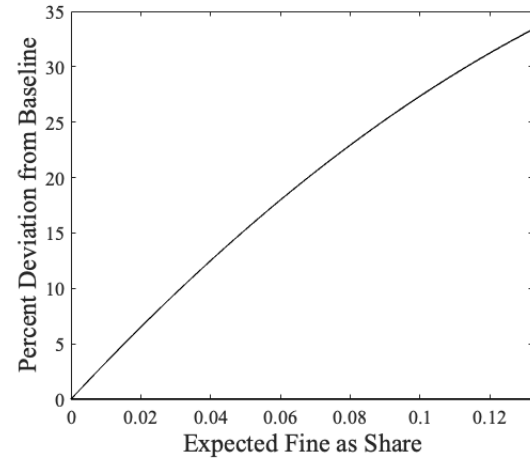


Figure 8: Effects of Policy Experiments on Vacancies by Sector

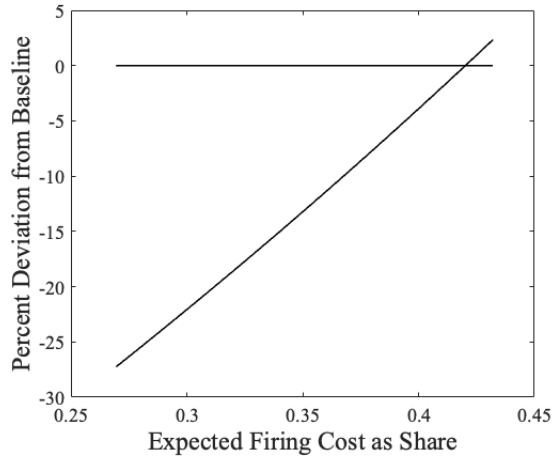
(a) Firing Costs: FS Vacancies



(b) Informality Penalty: FS Vacancies



(c) Firing Costs: IS Vacancies



(d) Informality Penalty: IS Vacancies

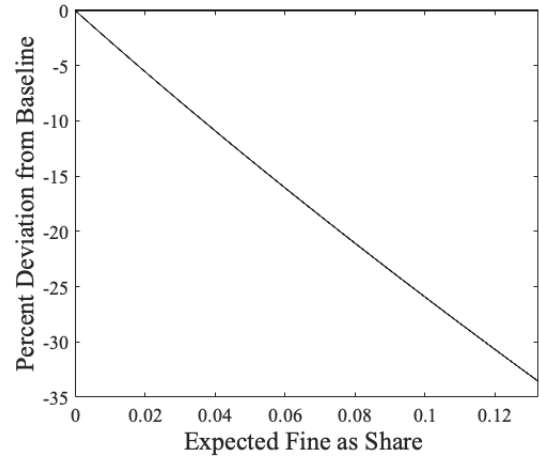
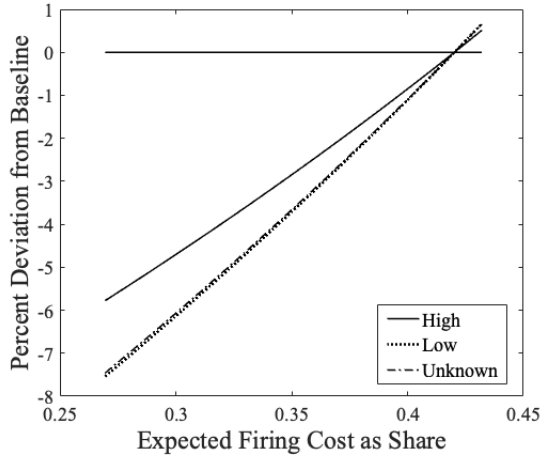
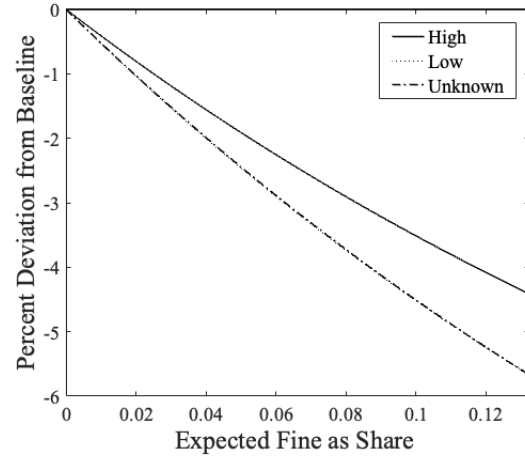


Figure 9: Effects of Policy Experiments on Cutoffs by Sector

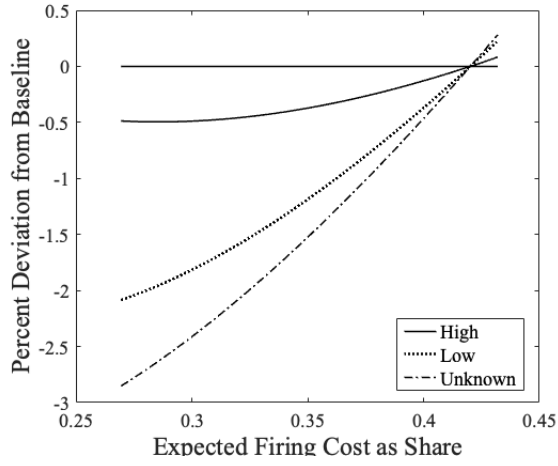
(a) Firing Costs: FS Cutoffs



(b) Informality Penalty: FS Cutoffs



(c) Firing Costs: IS Cutoffs



(d) Informality Penalty: IS Cutoffs

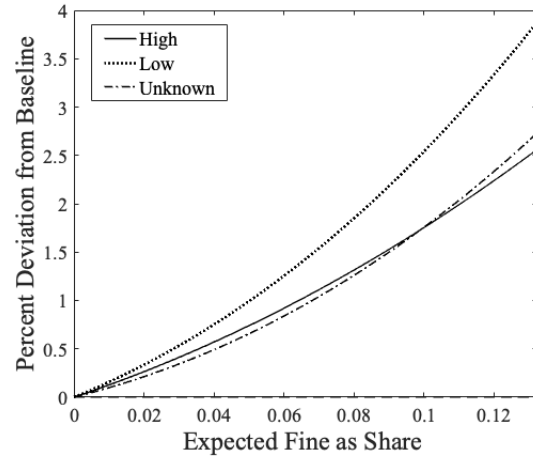
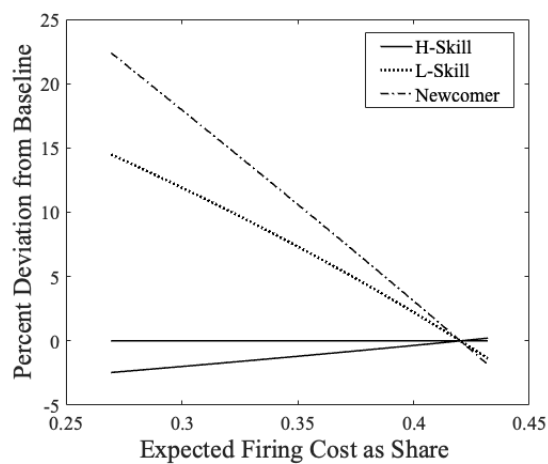
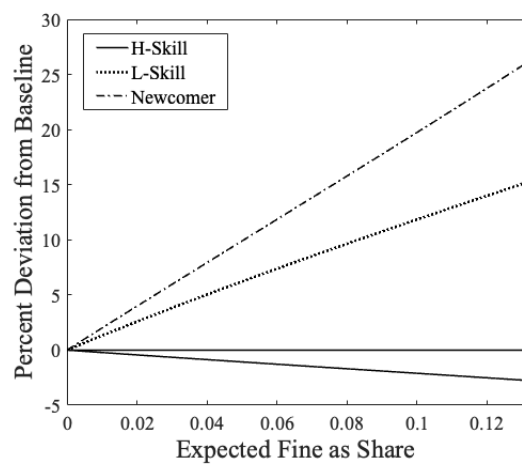


Figure 10: Effects of Policy Experiments on Employment Share by Skill and Sector

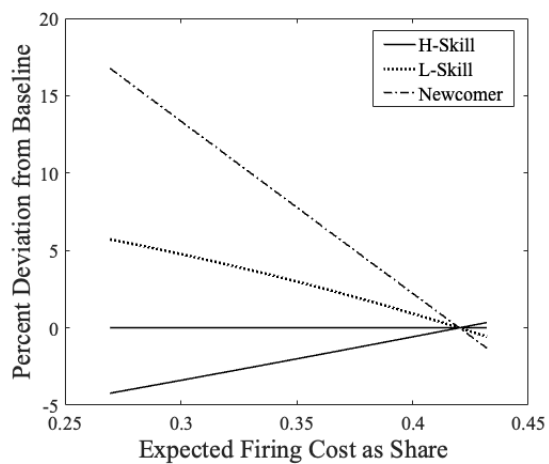
(a) Firing Costs: Employment FS



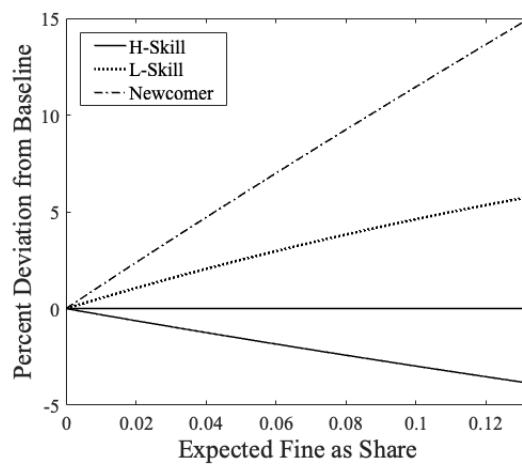
(b) Informality Penalty: Employment FS



(c) Firing Costs: Employment IS



(d) Informality Penalty: Employment IS



develop a two-sector labor market search model with heterogeneity in workers' initial skill levels. Firms post vacancies in both the formal and informal sectors, and workers are free to search for employment opportunities across both sectors. While all workers are born with an unknown level of skill, their skill-level is revealed probabilistically while working in either sector, a process we refer to as employer screening. Similarly, a low skilled worker will probabilistically gain human capital and become high skilled while working in either sector.

We use the ENOE, a household survey in Mexico that collects detailed income and employment data, to both estimate the rates of employer screening and human capital accumulation across employment sectors and to calibrate the remaining parameters of our model. Using our calibrated model, we find that hiring standards are substantially higher in the formal sector relative to the informal sector. Specifically, the hiring standards for new labor market entrants with unknown skill levels (newcomers) are found to be approximately 15% higher in the formal sector than in the informal sector. This large gap in hiring standards across sectors serves as evidence that the informal sector is operating as a “port-of-entry” into the labor market for young less-educated workers. The baseline calibration also reveals the important roles of the informal sector in solving an information problem about the workers' skills through employer screening and the opportunity to accumulate skills. While both sectors contribute to these processes, the informal sector plays an important role as it is the main employer of newcomers and low-skilled workers. Moreover, the survey data used in the analysis suggest that the informal sector seems to be more efficient in both roles for the case of young less-educated workers in Mexico.

We also consider a set of counterfactual policy experiments. While both policies intend to reduce the size of the informal sector, one policy achieves this by reducing firing costs in the formal sector (deregulation) while the other increases the punishment/fine faced by firms caught operating informally (cracking down). Our results favor deregulation over cracking down, as this policy simultaneously reduces the size of the informal sector and unemployment in equilibrium. However, both policies have the unintended effect of reducing the proportion of agents that are high-skilled and increasing the proportion that are either newcomers or low-skilled. This finding suggests that by serving as a “port-of-entry” into the labor market, informal positions play an important role in the skill revelation and accumulation processes for young less-educated workers. As such, policymakers who want to limit the scope of the informal sector should consider additional policies that help formal sector firms screen new labor market entrants and provide additional opportunities for less-educated workers to gain human capital.

References

- Amaral, P. S. and E. Quintin (2006). A competitive model of the informal sector. *Journal of Monetary Economics* 53, 1541–1553.
- Arias, O. S. and W. F. Maloney (2007). The *razón de ser* of the informal worker. In G. E. Perry, W. F. Maloney, O. S. Arias, P. Fajnzylber, A. D. Mason, and J. Saavedra-Chanduvi (Eds.), *Informality: Exit and exclusion*, Chapter 2, pp. 43–78. Washington, D.C.: The World Bank.
- Babington, M. and J. Cano-Urbina (2018). Stock sampling with interval-censored elapsed duration: A monte carlo analysis. Working paper, Florida State University.
- Becker, G. S. (1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education* (3d ed.). Chicago: The University of Chicago Press (for NBER).
- Bobba, M., L. Flabbi, S. Levy, and M. Tejada (2019, June). Labor market search, informality, and on-the-job human capital accumulation. TSE Working Paper 983, Toulouse School of Economics.
- Bosch, M. and J. Esteban-Pretel (2012). Job creation and job destruction in the presence of informal markets. *Journal of Development Economics* 98(2), 270 – 286.
- Burdett, K., C. Carrillo-Tudela, and M. G. Coles (2011). Human capital accumulation and labor market equilibrium. *International Economic Review* 52(3), 657–677.
- Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics : Methods and Applications*. Cambridge ; New York: Cambridge University Press.
- Cano-Urbina, J. (2015). The role of the informal sector in the early careers of less-educated workers. *Journal of Development Economics* 112, 33 – 55.
- Cano-Urbina, J. (2016, January). Informal labor markets and on-the-job training: Evidence from wage data. *Economic Inquiry* 54(1), 25–43.
- Card, D. and A. B. Krueger (2000). Minimum wages and employment: a case study of the fast-food industry in new jersey and pennsylvania: reply. *American Economic Review* 90(5), 1397–1420.
- Congress (1970). *Federal Labor Law*. Mexico, D.F.
- Dolado, J. J., M. Jansen, and J. F. Jimeno (2009, January). On-the-job search in a matching model with heterogeneous jobs and workers. *The Economic Journal* 119(534), 200–228.
- Engbom, N. (2022, January). Labor market fluidity and human capital accumulation. Working Paper 29698, National Bureau of Economic Research.
- Faberman, R. J., A. I. Mueller, A. Sahin, and G. Topa (2017, August). Job search behavior among the employed and non-employed. Working Paper 23731, National Bureau of Economic Research.
- Fields, G. S. (2009). Segmented labor market models in developing countries. In H. Kincaid and D. Ross (Eds.), *The Oxford Handbook of Philosophy of Economics*, Chapter 18, pp. 476–510. Oxford: Oxford University Press.

- Heckman, J., C. Pagés-Serra, A. C. Edwards, and P. Guidotti (2000). The cost of job security regulation: Evidence from Latin American labor markets [with comments]. *Economía* 1(1), 109–154.
- Hemmer, H.-R. and C. Mannel (1989). On the economic analysis of the urban informal sector. *World Development* 17(10), 1543–1552.
- Horvath, J. and G. Yang (2022). Unemployment dynamics and informality in small open economies. *European Economic Review* 141, 103949.
- INEGI (2005). *Encuesta Nacional de Ocupación y Empleo 2005. Una nueva encuesta para México. ENOE*. Aguascalientes: Instituto Nacional de Estadística, Geografía e Informática.
- INEGI (2007). *Conociendo la base de datos de la ENOE*. Aguascalientes: Instituto Nacional de Estadística, Geografía e Informática.
- Jovanovic, B. (1979, October). Job matching and the theory of turnover. *Journal of Political Economy* 87(5), 972–990.
- Lancaster, T. (1990). *The econometric analysis of transition data*. New York: Cambridge University Press.
- Levy, S. (2007, January). Can social programs reduce productivity and growth? A hypothesis for Mexico. Working Paper 37, International Policy Center, Gerald R. Ford School of Public Policy, University of Michigan.
- Loayza, N. V. (1996). The economics of the informal sector: a simple model and some empirical evidence from Latin America. *Carnegie-Rochester Conference Series on Public Policy* 45, 129–162.
- Montenegro, C. E. et al. (2007). Job security and the age-composition of employment: evidence from Chile. *Estudios de Economía* 34(2), 109–139.
- Montes Rojas, G. and M. Santamaría (2007). The burden of labour costs in Mexico. *LABOUR* 21(1), 157–188.
- Pissarides, C. A. (1994). Search unemployment with on-the-job search. *The Review of Economic Studies* 61(3), 457 – 475.
- Pissarides, C. A. (2000). *Equilibrium unemployment theory* (2nd ed.). Cambridge, Mass.: MIT Press.
- Samaniego de la Parra, B. and L. Fernandez Bujanda (2020, October). Increasing the cost of informal workers: Evidence from Mexico. Working paper, Banco de México.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: The MIT Press.

Technical Appendix: Assessing the Impact of Informal Sector Employment on Young Less-Educated Workers

Javier Cano-Urbina and John Gibson

In this appendix we provide details regarding the steady-state of our two sector labor search model. We also provide additional information regarding the estimation of sector-specific revelation and accumulation rates.

A Wages

Wages for all workers in the labor market are determined according to a surplus-sharing rule that entitles workers to a fraction β of the match surplus. In equilibrium, free entry implies that the profit from one more vacancy in the formal and the informal sectors is zero, and so in equilibrium it is the case that $V_F = V_I = 0$. Matches of a firm in sector $j \in \{F, I\}$ and a worker are given by $S_{jk}(x) = W_{jk}(x) - U_k + J_{jk}(x) - V_j$ for $k \in \{N, H, L\}$, and so the surplus-sharing rule dictates that $[W_{jk}(x) - U_k] = \beta S_{jk}(x)$. In equilibrium, free entry and the surplus sharing rule result in formal- and informal-sector wages for workers with known skill level given by:

$$(1a) \quad w_{FH}(x) = \beta p_{FH}x + (1 - \beta)(r + \tau)U_H - \beta \lambda_F D$$

$$(2a) \quad w_{FL}(x) = \beta p_{FL}x + (1 - \beta)(r + \tau)U_L - \beta \lambda_F D - (1 - \beta)\kappa_F[U_H - U_L]$$

$$(3a) \quad w_{IH}(x) = \beta p_{IH}x + (1 - \beta)(r + \tau)U_H - \beta \pi T - (1 - \beta)\eta f_F \int_{C_{FH}}^1 [W_{FH}(x) - U_H] dG(x)$$

$$(4a) \quad w_{IL}(x) = \beta p_{IL}x + (1 - \beta)(r + \tau)U_L - \beta \pi T - (1 - \beta)\eta f_F \int_{C_{FL}}^1 [W_{FL}(x') - U_L] dG(x') \\ - (1 - \beta)\kappa_I[U_H - U_L]$$

where all wage contracts are functions of the match quality. If the job is destroyed in the formal sector the firm incurs a firing cost D and since the firm and worker are sharing the match surplus the worker's wage in this sector is reduced by a fraction β of the expected firing cost. Similarly, if authorities catch an informal-sector job the firm incurs a penalty of T and given the surplus sharing rule the worker's wage is reduced by a fraction β of the expected penalty. A low-skilled worker that accumulates skills and become high skilled experiences a gain in lifetime utility of $[U_H - U_L]$; given the surplus sharing rule the worker shares this gain with the firm and so the wage for low-skilled workers is additionally reduced

by a fraction $(1 - \beta)$ of the gain. Finally, an informal-sector worker that moves to the formal sector experiences a gain of $E[W_{FL}(x') - U_k | x' > C_{Fk}]$ and this gain is shared with the firm, and so wages in the informal sector are reduced by the expected gain from such movement.

Similarly, free entry and the corresponding surplus-sharing rule result in wages for newcomers given by:

$$(5a) \quad w_{FN}(x) = \beta p_{FN}x + (1 - \beta)(r + \tau)U_N - \beta \lambda_F D - (1 - \beta)\sigma_F[\nu U_L + (1 - \nu)U_H - U_N] \\ - \beta \sigma_F \nu (1 - \Gamma_{FL}(x))D - \beta \sigma_F (1 - \nu)(1 - \Gamma_{FH}(x))D$$

$$(6a) \quad w_{IN}(x) = \beta p_{IN}x + (1 - \beta)(r + \tau)U_N - \beta \pi T - (1 - \beta)\eta f_F \int_{C_{FN}}^1 [W_{FN}(x) - U_N]dG(x) \\ - (1 - \beta)\sigma_I[\nu U_L + (1 - \nu)U_H - U_N]$$

where now wage contracts for newcomers account for the possible gain associated with having their skill level revealed given by $[\nu U_L + (1 - \nu)U_H - U]$. In the formal sector, wage contracts also account for the possibility that once the worker's skill level is revealed, the match might have to be destroyed and the firm incur firing cost, D . Wages in the formal sector for newcomers account for the possibility that the match may be exogenously destroyed before the worker's type is revealed. This possibility is also accounted for in the expression for the informal sector wage, but this wage must also account for the possibility the worker might quit and move to the formal sector before having their skill level revealed.

In this model, reservation wages w_{jk}^R for $j \in \{F, I\}$ and $k \in \{N, H, L\}$ are obtained when we substitute match quality x with the corresponding hiring standard C_{jk} . So that the reservation wage is given by:

$$w_{jk}^R = w_{jk}(C_{jk})$$

for $j \in \{F, I\}$ and $k \in \{N, H, L\}$, where $w_{jk}(\cdot)$ is given by the equations (1a)-(6a).

B Simplified (Closed-Form) Hiring Standards

Contacts between job seekers and firms in both sectors result in a match if and only if the match quality drawn when they make contact is higher than a reservation match quality. The reservation match quality depends on the sector of employment, and whether the worker is a newcomer, high-skilled, or low-skilled. The reservation match quality C_{jk} is such that $J_{jk}(C_{jk}) = V_j$, for $j \in \{F, I\}$ and $k \in \{N, H, L\}$. In equilibrium, free entry implies that

$V_j = 0$ and so the cutoffs for high-skilled workers C_{FH} and C_{IH} solve:

$$(1b) \quad p_{FH}C_{FH} = (r + \tau)U_H + \lambda_F D$$

$$(2b) \quad p_{IH}C_{IH} = (r + \tau)U_H + \pi T - \beta \frac{\eta f_F}{r + \tau + \lambda_F} p_{FH} [\hat{x}_{FH} - \bar{G}(C_{FH})C_{FH}]$$

where $\hat{x}_{FH} = \int_{C_{FH}}^1 x dG(x)$ is the average match quality in the formal sector for high-skilled workers.

For low-skilled workers C_{FL} and C_{IL} solve:

$$(3b) \quad p_{FL}C_{FL} = (r + \tau)U_L + \lambda_F D - \kappa_F[U_H - U_L] - \left(\frac{\kappa_F}{r + \tau + \lambda_F} \right) p_{FH}(C_{FL} - C_{FH})$$

$$(4b) \quad p_{IL}C_{IL} = (r + \tau)U_L + \pi T - \kappa_I[U_H - U_L] - \left(\frac{\kappa_I}{r + \tau + \lambda_I + \pi + \mu_{FH}} \right) p_{IH}(C_{IL} - C_{IH}) \\ - \beta \frac{\eta f_F}{r + \tau + \lambda_F + \kappa_F} \left(p_{FL} + \frac{\kappa_F}{r + \tau + \lambda_F} p_{FH} \right) [\hat{x}_{FL} - \bar{G}(C_{FL})C_{FL}]$$

where $\hat{x}_{FL} = \int_{C_{FL}}^1 x dG(x)$ is the average match quality in the formal sector for low-skilled workers.

The value of unemployment increases all measures of reservation match quality since unemployment is the outside option when considering taking a job. Note that firing costs and penalty costs increase the reservation match quality in the formal and informal sectors, respectively, for both high- and low-skilled workers. For low-skilled workers, the third and fourth terms in (3b) and (4b) indicate that the reservation match quality for these workers is reduced by the possibility of accumulating skills. Similarly, for informal-sector workers, the last terms in (2b) and (4b) indicate that the reservation match quality of these workers is reduced by the possibility of making a transition to the formal sector.

For newcomers, the reservation match qualities C_{FN} and C_{IN} solve:

$$(5b) \quad p_{FN}C_{FN} = (r + \tau)U_N + \lambda_F D - \sigma_F[\nu U_L + (1 - \nu)U_H - U_N] \\ - \Gamma_{FL}(C_{FN}) \frac{\sigma_F \nu}{r + \tau + \lambda_F + \kappa_F} \left(p_{FL} + \frac{\kappa_F}{r + \tau + \lambda_F} p_{FH} \right) (C_{FN} - C_{FL}) \\ - \Gamma_{FH}(C_{FN}) \frac{\sigma_F(1 - \nu)}{r + \tau + \lambda_F} p_{FH}(C_{FN} - C_{FH}) \\ + (1 - \Gamma_{FL}(C_{FN}))\nu\sigma_F D + (1 - \Gamma_{FH}(C_{FN}))(1 - \nu)\sigma_F D$$

$$\begin{aligned}
(6b) \quad p_{IN}C_{IN} &= (r + \tau)U_N + \pi T - \sigma_I[\nu U_L + (1 - \nu)U_H - U_N] \\
&\quad - \Gamma_{IL}(C_{IN}) \frac{\sigma_I \nu}{r + \tau + \lambda_I + \pi + \mu_{FL} + \kappa_I} \left(p_{IL} + \frac{\kappa_I}{r + \tau + \lambda_I + \pi + \mu_{FH}} p_{IH} \right) (C_{IN} - C_{IL}) \\
&\quad - \Gamma_{IH}(C_{IN}) \frac{\sigma_I(1 - \nu)}{r + \tau + \lambda_I + \pi + \mu_{FH}} p_{IH}(C_{IN} - C_{IH}) \\
&\quad - \eta f_F \int_{C_{FN}}^1 [W_{FN}(x') - U_N] dG(x')
\end{aligned}$$

As in the case of high- and low-skilled workers, the reservation match quality for newcomers depends on the value of unemployment, the firing costs, and the penalty costs. However, for these workers the reservation match quality also depends on the gains associated with the process of the revelation of skills. Gains associated with the revelation of skills reduce the reservation match quality and these are the gains for both the worker and the firm and are represented by the third, fourth, and fifth terms in equations (5b) and (6b). In the case of the formal sector, there are also costs associated with the revelation of skills. These are the costs incurred if the match has to be destroyed once the worker skill level is revealed and it is indicated in the last two terms in the equation (5b). In all cases, the gains and costs associated with the revelation of skills depend on the value of C_{jN} with respect to C_{jH} and C_{jL} as indicated by the indicator functions $\Gamma_{jk}(x)$ for $j \in \{F, I\}$ $k \in \{H, L\}$. Finally, the reservation match quality in the informal sector also depends on the gains associated with the possibility that the newcomer moves from the informal to the formal sector before the skill level is revealed. This gain is represented by the last term in equation (6b).

C Steady-State Worker Flows

The equations for the worker flows represent the last block of equations in the definition of equilibrium. Figure C.1 presents a diagram of the equilibrium worker flows in our model, with each box representing a unique state for a worker. There are nine states in the labor market which correspond to every combination of the workers' employment status (unemployed, formal sector, and informal sector) and skill level (newcomer, L-skilled, and H-skilled). To avoid cluttering the picture, we do not specify the specific transition probabilities between these states, but instead we discuss them in the text below. For the same reason, this picture does not show the workers who permanently exit the labor market, as this would simply be shown as arrows out of each box in the figure. The equilibrium flow equations are presented and discussed in detail below.

First, all workers enter the labor market as newcomers through unemployment, the far left side of the diagram in box U_N . While unemployed, newcomers search for jobs in both the

formal and informal sectors. Unemployed newcomers, in box U_N , can move out of this state in three ways. With probability $f_F \bar{G}(C_{FN})$ a newcomer gets a formal-sector job moving to box FS_N , where $\bar{G}(\cdot) = 1 - G(\cdot)$. Similarly, with probability $f_I \bar{G}(C_{IN})$ the newcomer finds an informal-sector job, moving to box IS_N . Finally, with probability τ the newcomer exits the labor market permanently (this arrow is not shown in the picture).

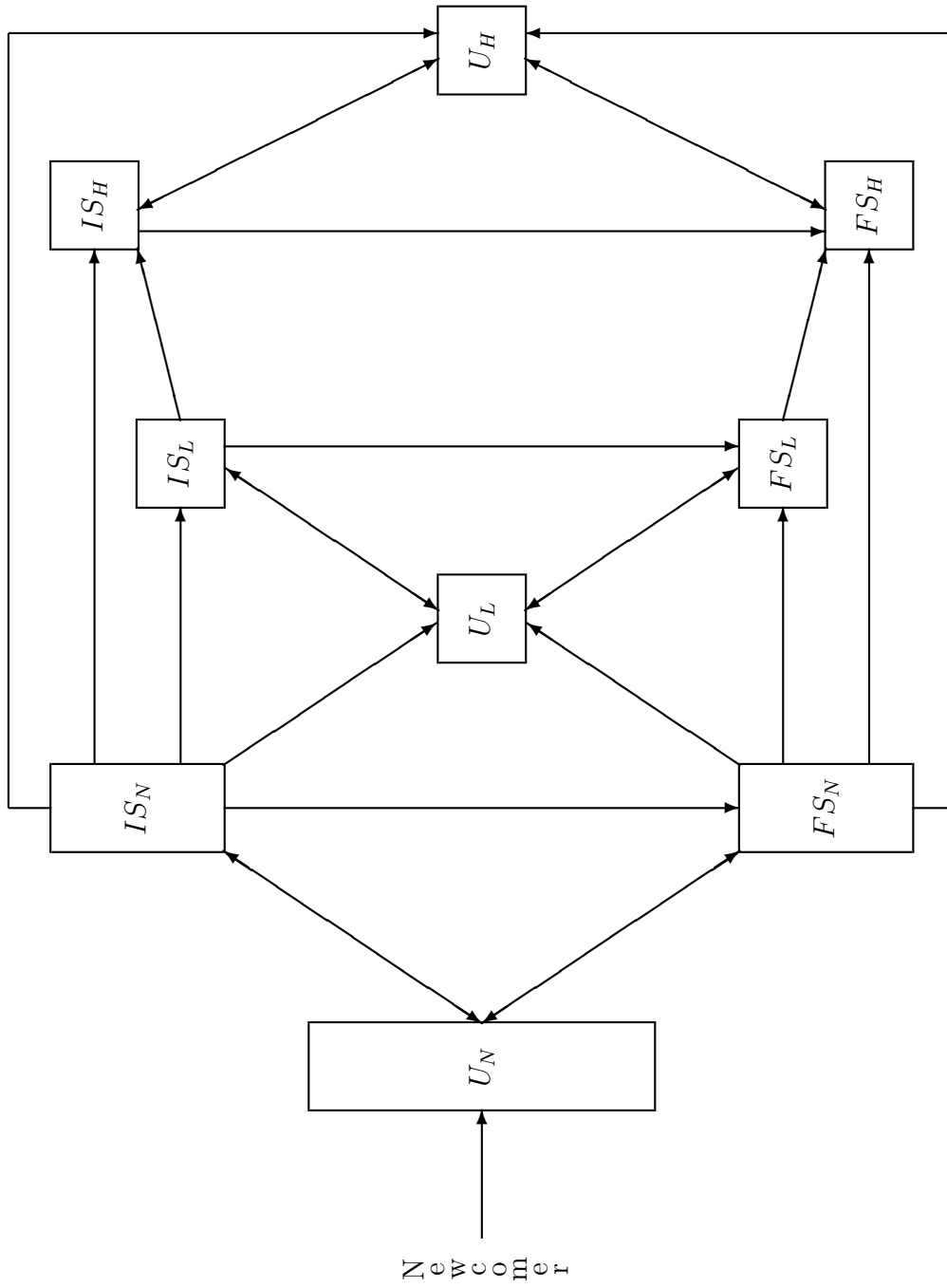
Newcomers employed in the formal sector, in box FS_N , can transition out of this state in three ways. First, they may lose their job due to exogenous job destruction with probability λ_F , which causes them to move back to box U_N . Second, they may have their type revealed with probability σ_F and move to boxes FS_H or FS_L , if the current match quality is good enough to keep the job, or to boxes U_H or U_L , if the match quality is insufficient to maintain the employment relationship. And lastly, the agent may exit the labor market permanently with probability τ (the arrow is not shown in the picture). Analogous transitions to the three described above apply to newcomers employed in the informal sector, in box IS_N . However, informal sector newcomers can transition out of this state through two additional channels: (i) with probability $\eta f_F \bar{G}(C_{FN})$ they get a formal-sector job moving to box FS_N , and (ii) with probability π the informal job is destroyed by authorities moving to box U_N .

Workers who are revealed as H-skilled while working in either the formal or informal sectors, keep moving between boxes FS_H , IS_H , and U_H until they permanently leave the labor market which happens with probability τ . Unemployed H-skilled workers in box U_H find a formal job with probability $f_F \bar{G}(C_{FH})$ and an informal job with probability $f_I \bar{G}(C_{IH})$ moving to boxes FS_H and IS_H , respectively. Once employed they can become unemployed if their job is exogenously destroyed with probability λ_F or λ_I , depending on sector of employment. Those workers employed in an informal job can also loose their job if detected by the authorities with probability π , and they can also move to the formal sector with probability $\eta f_F \bar{G}(C_{FH})$.

Workers who are revealed as L-skilled while working in either the formal or informal sector also keep moving between boxes FS_L , IS_L , and U_L . The transition probabilities between these three states are similar to those of H-skilled described in the previous paragraph, except that the match qualities must be now higher than the cutoff C_{jL} , $j \in \{F, I\}$. Similarly, L-skilled exit the labor market permanently with probability τ . Compared to H-skilled, L-skilled workers have an additional transition which is moving from boxes FS_L to FS_H or from IS_L to IS_H when they accumulate human capital and become H-skilled. Such workers accumulate skills at rates κ_F and κ_I , depending on their current sector of employment.

Next, we describe the equilibrium flows in equations (1c)-(9c), which are graphically represented in Figure C.1. In these equations, u_k , n_{Fk} , and n_{Ik} represent the number of workers who are unemployed, employed in the formal sector and employed in the informal

Figure C.1: Worker Flows



NOTES: Every box in the figure represents a particular worker's state. To avoid cluttering the figure even more we do not include the transition probabilities between these workers' states. These probabilities can be inferred from equations (1c)-(9c). Similarly, to avoid visual cluttering we do not include the probabilities of exiting the labor market permanently, which would be represented as an outward arrow from each box.

sector, respectively, as a fraction of the labor force, for skill-level $k \in \{N, H, L\}$.

$$(1c) \quad [\tau + f_F \bar{G}(C_{FN}) + f_I \bar{G}(C_{IN})]u_N = \lambda_F n_{FN} + \lambda_I n_{IN} + \pi n_{IN} + \tau$$

$$(2c) \quad [\tau + \lambda_F + \sigma_F]n_{FN} = f_F \bar{G}(C_{FN})u_N + \eta f_F \bar{G}(C_{FN})n_{IN}$$

$$(3c) \quad [\tau + \lambda_I + \eta f_F \bar{G}(C_{FN}) + \pi + \sigma_I]n_{IN} = f_I \bar{G}(C_{IN})u_N$$

$$(4c) \quad [\tau + \lambda_F + \kappa_F]n_{FL} = f_F \bar{G}(C_{FL})u_L + \eta f_F \bar{G}(C_{FL})n_{IL} + \sigma_F \nu \bar{G}(C_{FL}|x \geq C_{FN})n_{FN}$$

$$(5c) \quad [\tau + \lambda_I + \pi + \kappa_I + \eta f_F \bar{G}(C_{FL})]n_{IL} = f_I \bar{G}(C_{IL})u_L + \sigma_I \nu \bar{G}(C_{IL}|x \geq C_{IN})n_{IN}$$

$$(6c) \quad [\tau + \lambda_F]n_{FH} = f_F \bar{G}(C_{FH})u_H + \eta f_F \bar{G}(C_{FH})n_{IH} + \kappa_F n_{FL} + \sigma_F(1 - \nu)\bar{G}(C_{FH}|x \geq C_{FN})n_{FN}$$

$$(7c) \quad [\tau + \lambda_I + \pi + \eta f_F \bar{G}(C_{FH})]n_{IH} = f_I \bar{G}(C_{IH})u_H + \kappa_I n_{IL} + \sigma_I(1 - \nu)\bar{G}(C_{IH}|x \geq C_{IN})n_{IN}$$

$$(8c) \quad [\tau + f_I \bar{G}(C_{IH}) + f_F \bar{G}(C_{FH})]u_H = \\ = (\lambda_I + \pi)n_{IH} + \lambda_F n_{FH} + \sigma_F(1 - \nu)G(C_{FH}|x \geq C_{FN})n_{FN} + \sigma_I(1 - \nu)G(C_{IH}|x \geq C_{IN})n_{IN}$$

$$(9c) \quad [\tau + f_I \bar{G}(C_{IL}) + f_F \bar{G}(C_{FL})]u_L = \\ = (\lambda_I + \pi)n_{IL} + \lambda_F n_{FL} + \sigma_F \nu G(C_{FL}|x \geq C_{FN})n_{FN} + \sigma_I \nu G(C_{IL}|x \geq C_{IN})n_{IN}$$

where $\bar{G}(\cdot) = 1 - G(\cdot)$. Since all worker shares are represented as a fraction of the labor force we have that:

$$(10c) \quad u_N + n_{IN} + n_{FN} + u_L + n_{IL} + n_{FL} + u_H + n_{IH} + n_{FH} = 1$$

The left-hand side of the flows in equations (1c)-(9c) represent all the worker flows out of a given state and the right-hand side represent all the worker flows into that particular state. For example, equation (1c) is the equilibrium flow of newcomers into and out of unemployment. On the left-hand side we have the three channels out of the unemployment state for newcomers: they could exit the labor market permanently with probability τ , they could contact a formal-sector firm with probability f_F and with probability $\bar{G}(C_{FN})$ they create a match, or they could contact an informal-sector firm with probability f_I and with probability $\bar{G}(C_{IN})$ they create a match. Similarly, on the right-hand side we have the four channels into unemployment for newcomers. The first three channels into the newcomers' unemployment are: (i) a match can be exogenously destroyed while a worker is employed in the formal sector with probability λ_F , (ii) or destroyed with probability λ_I if employed in the informal sector, or (iii) destroyed with probability π if employed in the informal sector and the job is detected by authorities. The fourth channel into the unemployment state for newcomers comes from the replacement of workers who permanently exited the labor market

and so a fraction, τ , of all workers flow into unemployment when they enter the labor market for the first time. All of these transitions are depicted in Figure C.1, with the exception of the transitions out of the labor market which would be depicted as arrows pointing out from each of the nine boxes in the picture.

Similarly, equations (2c) and (3c) show that workers can leave the state of newcomer employment in the formal or informal sector, respectively, as either L- or H-skilled with probabilities σ_F and σ_I , respectively. Notice that equations (4c)-(9c) describe that when workers' skills are revealed, these workers can loose their jobs. We do not make any assumption about the relative size of the cutoffs C_{jN} , C_{jH} , and C_{jL} , for $j \in \{F, I\}$. Hence, it is possible that a newcomer is hired in the formal sector with match quality $\tilde{x}$ larger than C_{FN} , but that once the worker's skill level is revealed, the match quality $\tilde{x}$ is not larger than the hiring standard corresponding to his revealed type in the formal sector. Equations (4c) and (6c) indicate that only formal-sector workers with a match quality larger than the cutoff C_{Fk} for $k \in \{H, L\}$ would keep their job and renegotiate their wage with the firm given their revealed skill level. Those whose current match quality is lower than the cutoff would become unemployed and join the group of unemployed of skill level $k \in \{H, L\}$ as described in equations (8c) and (9c). Something similar occurs when skills are revealed in the informal sector as indicated by equations (5c) and (7c) for workers that keep their informal-sector job and by equations (8c) and (9c) for workers that loose their informal-sector job.

Finally, these equations show the flow of L-skilled workers as they accumulate human capital and become H-skilled. These transitions are illustrated in equations (4c) and (6c) for workers who accumulate skills while employed in the formal sector. Similarly, equations (5c) and (7c) illustrate the worker transitions when low-skilled workers accumulate skills while employed in an informal-sector job.

D Summary Statistics

Table D.1 presents the summary statistics for the sample. As the table indicates formal sector workers are on average older, more educated, more likely to be married, and have higher hourly earnings than informal sector workers. Also formal sector workers tend to be concentrated in larger firms whereas informal sector workers tend to be concentrated in small firms. In terms of the industry of occupation, formal sector workers are mainly concentrated in manufacturing and services whereas informal sector workers tend to be concentrated in services and construction.

Table D.2 presents the quarterly transition probabilities between informal employment, formal employment, and unemployment. The last elements in the first and second rows

Table D.1: Summary Statistics ENOE Q1:2005 - Q4:2018

	Formal Sector	Informal Sector
Age	24.35	22.97
Education	8.74	7.98
Married	0.31	0.19
Hourly earnings	28.97	25.19
Weekly hours worked	50.67	49.55
Firm Size (%)		
1-5	7.46	67.33
6-50	39.52	28.31
51+	53.02	4.37
Industry (%)		
Construction	8.55	28.44
Manufacturing	36.49	20.06
Commerce	22.96	15.72
Services	32.00	35.77
Sample size	144,000	132,755

NOTES: Male with less than 12 years of education and not attending school, ages 16 to 30. Individual and job characteristics at the time of the first interview. Hourly earnings are in Mexican pesos as in the second half of July of 2018 (in this same period the exchange rate was on average 18.82 Mexican Pesos for 1 US Dollar). The sample size for some of the statistics in the table are smaller due to missing values. In particular, for hourly earnings and weekly hours worked ($n_{FS} = 114,658, n_{IS} = 107,084$), firm size ($n_{FS} = 132,360, n_{IS} = 127,790$), and industry ($n_{FS} = 141,611, n_{IS} = 128,926$).

Table D.2: Quarterly Transition Probabilities ENOE Q1:2005 - Q4:2018

	Informal Sector	Formal Sector	Unemployed
Informal Sector	0.765	0.164	0.071
Formal Sector	0.134	0.823	0.043
Unemployment	0.387	0.255	0.358

of Table D.2 determine the flow of workers transitioning into unemployment from informal and formal employment, respectively. Inspection of Table D.2 shows that this separation rate is higher for the informal sector than the formal sector. Also, inspection of the last row of Table D.2 shows that unemployed individuals are more likely to flow into informal employment than formal employment.

E Estimating Employer Learning and Human Capital Accumulation Rates

E.1 Duration of Wage Changes Data from the ENOE

We construct our measures of duration of wage changes in three steps. The first step is to purge wages from individual observable characteristics. With this objective in mind, we estimate two separate regressions for each sector of employment assuming that the wage distribution is described by a Mincer's earnings function:

$$\ln w_i = x_i' \beta + \varepsilon_i$$

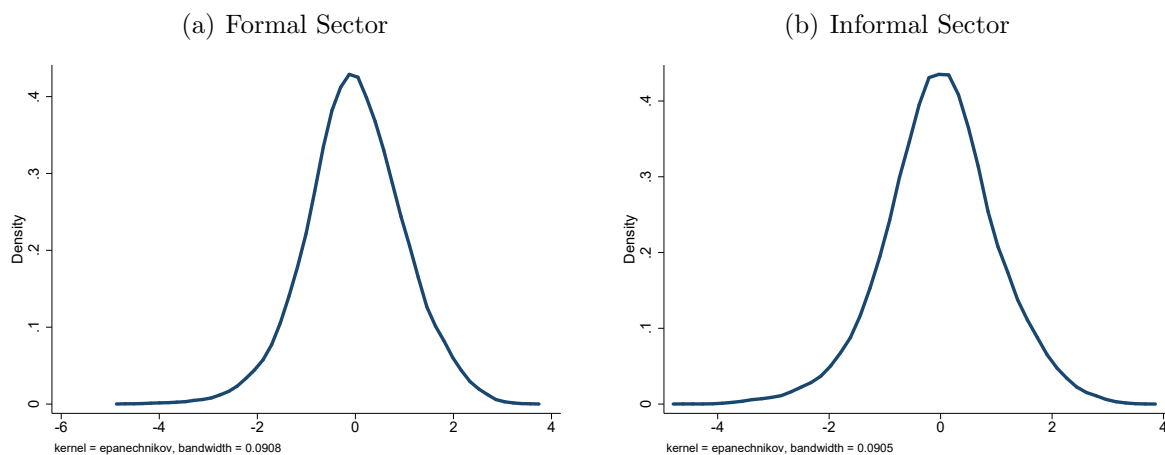
where w_i represent the hourly earnings and x_i is a vector of observable individual characteristics that include education, age, firm size, industry of occupation, controls for minimum-wage zones in Mexico as well as year and state dummies. Our sample include all males ages 16-30 with less than 12 years of education. We drop the top and bottom 0.5% of hourly earnings to reduce the effect of outliers.

These regressions are done separately for the informal and formal sectors and the results are presented in Table E.1 where it can be seen that all the human capital covariates have the expected signs with more education leading to higher earnings and with age and age-squared capturing the increasing returns to labor market experience at a decreasing rate. The firm-size controls indicate that larger firms in the informal sector pay on average higher wages than 1-5 employees firms, which is the omitted category. The firm-size controls are

not statistically significant for formal-sector workers which could be the result of not having many workers employed in very small firms in the formal sector, as indicated in the summary statistics in Table D.1. The industry of occupation has a similar effect on hourly earnings for the formal and informal sectors where all three Manufacturing, Services and Construction pay on average higher wages than Commerce, which is the omitted category. As indicated in Table D.1, informal-sector workers are predominantly employed in small firms in the Services and Construction industries whereas formal-sector workers are predominantly employed in large firms in the Manufacturing and Services industries.

The second step involves obtaining the distribution of wages purged from the individual observable characteristics. We do this by standardizing the residuals from the regressions presented in Table E.1. The standardized wage measure is then used to determine the wage distribution in each sector. These distributions are presented in Figure E.1. These distributions are then divided into deciles to determine the samples used to estimate σ_j and κ_j , for $j \in \{F, I\}$. The two separate samples to estimate σ_F and σ_I are composed of all individuals whose standardized wage at the first interview is in the middle of the corresponding distribution of standardized wages (formal or informal). Specifically, those individuals whose standardized wage at the first interview is in the fifth or sixth deciles. The two separate samples to estimate κ_F and κ_I are composed of all individuals whose standardized wage at the first interview is in the bottom of the corresponding distribution, formal or informal distribution of standardized wages. Specifically, those individuals whose standardized wage at the first interview is in the first or second deciles.

Figure E.1: Distribution of Standardized Wage Measures by Sector



Notes: Data from the *Encuesta Nacional de Ocupacion y Empleo* (ENOE) from first quarter of 2005 to fourth quarter of 2018. Standardized measures of the residuals from the log-wage regressions presented in Table E.1.

The third step involves identifying the point in time in which the individuals in each of

Table E.1: Log-Wage Regressions by Sector

	Formal Sector	Informal Sector
Graduate grade 6	0.0345*** (0.0075)	0.0369*** (0.0060)
Graduate grade 9	0.0825*** (0.0072)	0.0651*** (0.0058)
Age	0.0798*** (0.0044)	0.0982*** (0.0042)
Age squared	-0.0014*** (0.0001)	-0.0018*** (0.0001)
Married	0.0527*** (0.0028)	0.0744*** (0.0037)
Firm size		
6 - 20	-0.0007 (0.0049)	0.0760*** (0.0035)
21+	-0.0046 (0.0046)	0.0855*** (0.0051)
Industry		
Manufacturing	0.0718*** (0.0035)	0.0922*** (0.0049)
Services	0.0623*** (0.0034)	0.0631*** (0.0043)
Construction	0.2421*** (0.0049)	0.2375*** (0.0045)
No. of observations	72,100	63,115
R^2	0.1760	0.1926

Notes: Data from the *Encuesta Nacional de Ocupacion y Empleo* (ENOE) from first quarter of 2005 to fourth quarter of 2018. The sample includes males ages 16-30 with less than 12 years of education. We drop observations with wages above the top 0.5 percent or below the bottom 0.5 percent of the distribution of wages. The omitted firm size category is 1-5 employees. The omitted industry category is Commerce. The educational dummies are: Graduate Grade 6 = $\mathbf{1}\{Edu \geq 6\}$ and Graduate Grade 9 = $\mathbf{1}\{Edu \geq 9\}$. The covariates also include year dummies (2005-2018), state dummies, and dummies for the three minimum-wage zones in Mexico. Standard errors in parentheses.

*Significant at 10%, **Significant at 5%, ***Significant at 1%.

the four samples make a transition to a decile in the wage distribution that can be attributed to the revelation or the accumulation of skills. In the survival analysis literature, these are known as *failures*. Failures attributed to the revelation of skills or skill accumulation are done separately.

First, to identify failures than can be attributed to the revelation of skills we follow individuals who started in the fifth or sixth decile from the second through the fifth interview and locate the interview in which the standardized wage of these individuals is located in the first or the tenth deciles. We take this as an indication that the wage has been revised down for those revealed as L-skilled (in the first decile) and revised up for those revealed as H-skilled (in the tenth decile). The time between the first interview and the interview in which the standardized wage is in the first or tenth decile is our measure of the duration of a wage change used to estimate σ_F and σ_I .

Next, to identify failures that can be attributed to the accumulation of skills we follow individuals who started in the first or second decile from the second through the fifth interview and locate the interview in which the standardized wage of these individuals is located in the ninth or tenth deciles. We take this as an indication that the wage has been revised up after those L-skilled individuals accumulated enough skills to become H-skilled. The time between the first interview and the interview in which the standardized wage is in the ninth or tenth decile is our measure of the duration of a wage change used to estimate κ_F and κ_I .

Finally we impose further restrictions in our samples used to estimate σ_F , σ_I , κ_F , and κ_I . First, we restrict both samples to individuals who did not switch between the formal and informal sectors. This is because there are structural differences between the two sectors and so a worker who switches sectors might experience a change in wages that could be interpreted as a failure. Moreover, the cutoffs of the deciles are different for the standardized-wage distribution on each sector. Second, since we need to observe the wage we restrict the sample to individuals who have a valid wage measure during all the time they are in the sample. In both samples, this means having a valid wage measure during the five interviews of the ENOE. Third, in the samples to estimate σ_F and σ_I , we also include those workers who made a transition to unemployment and were re-employed in the same sector with a wage that located them in the first decile of the corresponding standardized-wage distribution. While these workers do not have five valid wage measures they are still included to account for those workers who lost their job after being revealed as L-skilled and did not have a sufficiently high match quality. Appendix E.4 provides more detail on the determination of failures from transitions to unemployment.

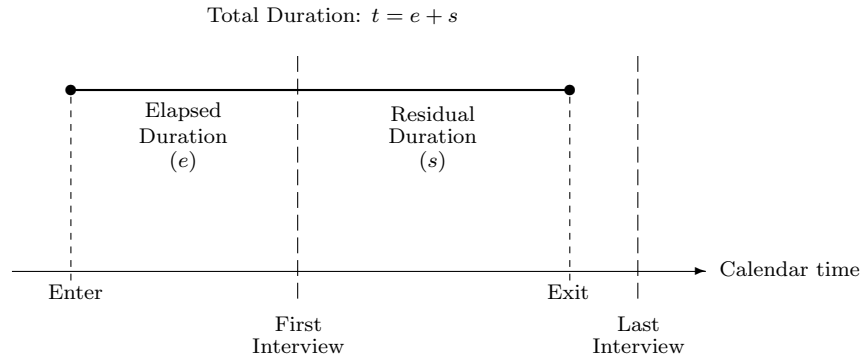
Individuals in each sample that never made a transition to a decile that can be attributed as a failure are right-censored. That is, for these individuals, all we know is that their

duration of wage changes is larger than 12 months, which is the time between the first and the last ENOE interviews. In the next section we describe more details about these duration measures and how are they used to estimate σ_F , σ_I , κ_F , and κ_I .

E.2 Estimating the Hazard Functions

In all four samples, the duration from first interview to the time when an individual experiences a wage change attributed to screening or human capital accumulation (a failure) is only part of the total duration within the state of interest. Given the way we build our duration measures, we are dealing with what is known as stock-sampling duration (see Lancaster, 1990; Wooldridge, 2002). As indicated on Figure E.2 when building duration measures from a stock of individuals, the total duration is composed of two parts: the elapsed duration and the residual duration. The residual duration is the length of time from the point when we start following the individual (e.g., the first interview) until the point when the individual leaves the state of interest. Elapsed duration is the time that passes between the individual entering the state of interest (e.g., becoming employed) and the point in time when we start following the individual (e.g., the first interview). The total duration is then the sum of elapsed and residual duration.

Figure E.2: Stock Sampling Duration



Given the stock-sampling scheme, the measures of duration described above for each sub-sample represent measures of residual duration. In both sub-samples, we only include individuals who held a job in the formal or informal sectors and did not switch sectors during the five interviews in the ENOE. The residual duration is measured in months and all residual duration measures are interval censored due to the fact that the ENOE collects household information at a quarterly frequency. That is, all durations will be contained in three-month intervals, $(0, 3]$, $(3, 6]$, $(6, 9]$, $(9, 12]$. The residual durations of individuals who did not experience a wage change that could be attributed to screening or human capital

accumulation during all five interviews are right censored. That is, for these individuals all we know is that their duration is in the interval $(12, \infty)$.

The next step in building the duration measures is to compute the elapsed duration. However, we do not observe the point in time when an individual becomes employed in the ENOE.³¹ Hence we are faced with a problem known in the duration analysis literature as *left-censoring* (Wooldridge, 2002). The direct approach to deal with left-censoring is to take a stand on the stochastic process that governs the elapsed durations (see Wooldridge, 2002, exercise 20.8), and then integrate out the elapsed duration from the likelihood function so that it only depends on the residual duration. While simply explained, this step is non-trivial.

In general, ignoring elapsed duration in the likelihood function and using only the residual duration leads to biased estimates of the parameters of the hazard function. However, there is a single scenario in which this is not the case and this is when the duration data follows an exponential distribution (see Example 3 in page 93 of Lancaster, 1990).³² This is an important observation for our purposes because in the model, both the revelation of a worker's skill level and the accumulation of skills are Poisson processes, which in turn implies that our duration data have an exponential distribution.

Define $T > 0$ as the random variable measuring duration of wage changes that are the result of employer screening or human capital accumulation. The exponential hazard function is a constant function of the form $\phi(t) = \phi$. When we estimate the hazard function with the samples of workers who start in the middle of the distribution in sector j then $\phi(t) = \sigma_j$, for $j \in \{F, I\}$. Similarly, when we estimate the hazard function with the samples of workers who start in the bottom of the distribution in sector j then $\phi(t) = \kappa_j$, for $j \in \{F, I\}$. Since all samples use the standardized measures of wages that are already purged from the effect of observable individual characteristics, both of the hazard functions we estimate do not include covariates. The likelihood function is then:

$$L(\phi) = \prod_{i=1}^n (\phi e^{-\phi t_i})^{d_i} (e^{-\phi t_i})^{1-d_i}$$

where d_i is an indicator that equals one for completed durations, those that made a transition to the corresponding deciles, and zero for right-censored durations. In this likelihood, $\phi = \sigma_j$ when restricting the samples to individuals that start in the middle of the distribution in

³¹There are ways of obtaining this information from the ENOE and imputing the elapsed duration of individuals (see Cano-Urbina, 2015; Babington and Cano-Urbina, 2018). However, this would substantially reduce our sample size. More importantly, given that we estimate an exponential hazard function (see below), this information does not improve the estimation of the parameters of interest.

³²Lancaster shows that when “completed durations are Exponentially distributed, that is, when we are observing a Poisson process at a fixed point on the time axis, then elapsed durations and the completed durations of entrants are identically distributed” (Lancaster, 1990, page 93).

sector j and $\phi = \kappa_j$ when restricting the sample to individuals that start in the bottom of distribution in sector j , for $j \in \{F, I\}$.

E.3 Verifying Exponential Hazard Function

As described in the previous section, we are using left-censored data to estimate an exponential hazard function. In this section, we explore whether the exponential hazard function is an appropriate assumption for the processes of screening and human capital accumulation described in our model developed in Section 2. As is well-known in the duration analysis literature, the exponential distribution is a special case of the Weibull distribution. The Weibull distribution has a hazard function given by:

$$(1e) \quad \phi(t) = \phi \alpha t^{\alpha-1}$$

where α is known as the *duration-dependence* parameter. When $\alpha < 1$ the hazard function is decreasing in the duration variable t , when $\alpha > 1$ is increasing in t , and when $\alpha = 1$ is constant. Therefore, when $\alpha = 1$ the Weibull hazard function reduces to the exponential hazard function with $\phi(t) = \phi$. This suggest a procedure to verify if our assumption about the screening and human capital accumulation processes being Poisson processes is supported in the data. That is, estimate a Weibull hazard function and verify that the estimate of α is equal to one.

Before we proceed and estimate the Weibull hazard function we need to remember that our duration data is left-censored. Unlike the estimation of the exponential hazard function, the estimation of the Weibull hazard is impacted by left-censoring. One way to deal with left-censoring is to follow Wooldridge and integrate out the unobserved elapsed duration (see Wooldridge, 2002, exercise 20.8). We follow a simpler approach that resembles this procedure by imputing the unobserved elapsed duration and then building measures of completed duration by adding the imputed elapsed duration and the observed residual duration (see Figure E.2 for reference to residual and elapsed durations). To impute the elapsed duration we assume that this duration follows some distribution and we randomly sample measures of elapsed duration from this distribution. We do this for each individual in our sample and estimate a Weibull hazard function where we are particularly interested in the estimate of α . Then we repeat this process 1,000 times and calculate the average value of α and its standard deviation for the different replications.

Table E.2 presents the results of this exercise. The table presents the average estimate of the duration-dependence parameter α together with its standard deviation over the 1,000 replications for the four subsamples we use to estimate σ_F , σ_I , κ_F , and κ_I . In all four

Table E.2: Estimate of Weibull Hazard Function

	Sample to Estimate	Sample to Estimate	Sample to Estimate	Sample to Estimate
	σ_F	σ_I	κ_F	κ_I
Average estimate of α	1.0417	1.0450	1.0896	0.9930
Standard deviation	0.0693	0.0792	0.0846	0.0938
Sample size	1,042	609	1,057	595

Notes: Data from the *Encuesta Nacional de Ocupacion y Empleo* (ENOE) from first quarter of 2005 to fourth quarter of 2018. Duration data obtained from transitions in the distribution of standardized measures of the residuals from the log-wage regressions presented in Table E.1.

subsamples the average estimate of the duration dependence parameter α is very close to 1 and in all four cases the 95 percent confidence interval clearly includes $\alpha = 1$. As a result, we do not find strong evidence in the data against our assumption that the skill revelation and skill accumulation processes in our model are Poisson processes.

E.4 Calculation of ν

We calculate the fraction of L-skilled workers, ν , using a back-of-the envelope procedure with the same data used to calculate σ_F and σ_I . That is, individuals age 16-30 who are in the middle of the distribution of wages at the time of the first interview. In Section 5 and in Appendix E.1, we defined a failure as the event in which an individual who started in the middle of the distribution: (i) moves to the lower or upper tails of the distribution, or (ii) moves to unemployment and is re-employed in the same sector with a wage located in the first decile. To proceed define the two fractions:

- s_{Hj} = fraction of failures moving from middle to top of distribution from one interview to the next in sector j
- s_{Lj} = fraction of failures moving from middle to bottom of distribution from one interview to the next or after an unemployment spell in sector j

where $j \in \{F, I\}$. Taking the fact that the ENOE is at the quarterly frequency, these two fractions represent the number of failures accumulated in three months. The model predicts that every quarter the fraction of newcomers revealed as L-skilled and H-skilled in the formal and informal sectors are:

$$s_{Lj} = (3 \times \sigma_j) \times \nu$$

$$s_{Hj} = (3 \times \sigma_j) \times (1 - \nu)$$

for $j \in \{F, I\}$ and where $(3 \times \sigma_j)$ reflects the fact that the data is at the quarterly frequency and that σ_j is at the monthly frequency. Notice that what we observe in the data for s_{Lj} is biased since some of the newcomers revealed as L-skilled will lose their job and we might not capture their re-employment. While we try to remedy this problem as explained above it is not fully resolved. As a result it is more reliable to work with s_{Hj} to pin down the value of ν as we expect that all newcomers revealed as H-skilled will keep their jobs. Then we can pin the value of ν as:

$$\nu = 1 - \frac{s_{Hj}}{3 \times \sigma_j}$$

for $j \in \{F, I\}$. This last identity makes clear that the value we pin down for ν will depend on whether we use the sample from the formal or informal sectors. Since ν is an economy-wide parameter we proceed by pooling the formal and informal sector and estimating a pooled σ following the exact same duration data used to estimate σ_F and σ_I but pooling all durations. When we do this we get an estimate of $\hat{\sigma} = 0.0467$ with standard error 0.0018, and then calculate the value of ν using this pooled σ in place of the sector-specific rates. Then we calculate ν as:

$$(2e) \quad \nu = 1 - \frac{s_H}{3 \times \sigma}$$

where s_H pools the fraction of failures moving from middle to top of distribution in both the formal and informal sectors. We should mention that the values of ν we obtain using the sector specific transitions and σ_j are very similar. We discuss these below once we obtain the value of ν .

The transitions to the top and bottom deciles are described in Table E.3 where we pooled transitions in both the formal and informal sectors. Panel A of Table E.3 presents the number of workers who make transitions from the middle to the top or bottom of the wage distribution after wages have been purged for the observable characteristics.³³ The first column indicates the number of workers at risk, that is those workers in the middle of the distribution who have not made a transition to the top or bottom of the wage distribution. These workers represent the newcomers in our theoretical model. The second, third, and fourth columns indicate the number of workers who made transitions to the top or bottom of distribution, that is the failures. The second column presents all transitions, the third presents the number of transitions to the top, the fourth to the bottom of the distribution.

Panel B of Table E.3 presents these transitions as a fraction of the workers who were

³³The covariates are the same as those used in the estimation of σ_F , σ_I , κ_F , and κ_I described in Section 5.

still at risk at the beginning of the quarter. Since the fraction of transitions varies slightly from the second to the fifth interviews, we use the average of these transitions. Then using the average in the third column as our approximation of s_H , that is $s_H = 0.0671$, and the estimate of pooled σ given above we use the calculation suggested in (2e) and get a value of $\nu = 0.5203$. This is the number we use in our calibration exercises.

As indicated above, we get very similar values of ν if we use sector-specific transitions s_{Hj} and sector specific employer learning rates σ_j . If we use formal-sector transitions and our estimate of σ_F we calculate a value of $\nu = 0.4914$ with an initial number of 1,042 individuals at risk (our initial sample size). If we use informal-sector transitions and our estimate of σ_I we get $\nu = 0.5626$ with an initial number of 609 individuals at risk. Notice that if we use these numbers of individuals at risk to get a weighted average of the ν we calculate on each specific case we get a weighted average of 0.5176 which is very similar to the $\nu = 0.5203$ we found above. This should alleviate any concerns about the sensitivity of using either the data from the formal or informal sector, or the pooled data from both sectors.

Table E.3: Transitions from Middle to Top and Bottom of Wage Distribution

Interview	At Risk	All Failures	H-Skill Failures	L-skilled Failures
A. Number of Workers				
2nd	1,651	244	127	117
3rd	1,407	184	93	91
4th	1,223	156	79	77
5th	1,067	120	65	55
B. Fraction from At Risk				
2nd		0.1478	0.0769	0.0709
3rd		0.1308	0.0661	0.0647
4th		0.1276	0.0646	0.0630
5th		0.1125	0.0609	0.0515
Average		0.1296	0.0671	0.0625