Frictional Labor Markets, Education Choices and Wage Inequality *

Manuel Macera † Hitoshi Tsujiyama ‡

February 2017

Abstract

This paper studies how education choices and labor market frictions interact in shaping wage inequality. We first observe that the wage premium of college graduates relative to high school graduates (between-group inequality) has almost tripled over the past thirty years in the U.S., whereas the wage dispersion measured as the variance of log wages conditional on educational attainments (within-group inequality) has also become one and half times larger for both education groups. We then construct a model of schooling investment and labor market frictions that generates supply and demand of skills and frictional wage differentials as equilibrium objects, which is essential for analyzing sources of inequality. The model offers a structural interpretation of the trends in both the wage premium and wage dispersion, providing a novel way to decompose these measures into variations driven by returns to education, worker and firm heterogeneity, and matching frictions. Calibrating the model to the U.S. data, we find that the increase in the wage premium is largely explained by the changes in worker heterogeneity induced by worker sorting via college education rather than returns to college. The increase in the wage dispersion is generally accounted for by changes in firm heterogeneity, whereas changes in worker heterogeneity also play an important role, especially for college workers.

*For useful comments and discussions we thank Naoki Aizawa, Cristina Arellano, Nicola Fuchs-Schündeln, Jonathan Heathcote, Loukas Karabarbounis, Leo Kaas, John Kennan, Philipp Kircher, Alex Ludwig, Svetlana Pashchenko, Ponpoje Porapakkarmas, Fabrizio Perri, Ctirad Slavik, Naoki Takayama, Satoshi Tanaka, as well as participants at many seminars and conferences. All errors remain our own.

†Colorado State University. macera@colostate.edu

‡Goethe University Frankfurt. hitoshi.tsujiyama@hof.uni-frankfurt.de

1. Introduction

Economists have long been interested in decomposing wage inequality into the variation driven by workers' characteristics, and the residual component associated with labor market uncertainty. This decomposition is of fundamental importance in policy making because it provides guidance with respect to effective ways of reducing wage inequality. Although observable worker characteristics have shown to be of limited success in explaining the overall variation in wages, in recent years the empirical literature has made substantial progress in studying the sources of inequality thanks to the availability of longitudinal matched employer-employee datasets. It is now well known that unobserved worker heterogeneity and labor market uncertainty in the form of between firm wage differentials help to reduce considerably the unexplained variation in wages[1]. Moreover, in the case of the U.S., Song et al. (2015) find that most of the increase in wage inequality in the last forty years is likely to be driven by the increase in labor market sorting.

Despite the progress in the empirical side, there are not many models that can offer a structural interpretation to its findings. In fact, the theoretical literature has typically addressed wage inequality in two different ways. The macro literature has focused on the role of education and human capital investment in shaping this inequality, treating the residual component of it as exogenous. In contrast, recent developments in search theory treat the residual component as an endogenous object, assuming however that the distribution of skills at the moment of labor market entry is exogenously given. The main goal of this paper is to integrate these approaches and provide a benchmark for interpreting the empirical evidence. To this end, we develop a model of schooling investment and labor market search that generates both a distribution of skills and residual wages as equilibrium objects. We use the model to provide a novel decomposition of wage inequality and shed light on how educational choices and labor market uncertainty interact in shaping it.

Our model divides the lifetime of a worker in two stages, the schooling stage and the labor market stage. In the schooling stage, workers face an optimal stopping problem in which they must decide whether to complete college education or to dropout from it. Education is modeled as a process that not only enhances workers' productivity but also sorts them into the labor market according to their educational attainment. In the labor market stage, workers and firms are matched randomly, bargain over wages and workers are allowed to search on-the-job, much

[1] Unobserved worker heterogeneity and between-firm differentials correspond to the worker and firm fixed effects in wage equations estimated using the methodology introduced by Abowd et al. (1999). Two important recent contributions in this line of research are Card et al. (2013), which applies this methodology to decompose wage inequality using German data, and Song et al. (2015), which uses confidential Social Security Administration W-2 data to decompose changes in the unconditional variance of log wages in the U.S.

in the spirit of Cahuc et al. (2006). Importantly, markets are segmented by level of schooling, which implies that job finding rates will depend on the distribution of educational attainment. Consequently, the equilibrium in the labor market is affected by the workers' choices in the schooling stage. In turn, workers' schooling choices will depend on labor market equilibrium via the value of being unemployed in each submarket. The interaction between the heterogeneity induced by optimal choices in the schooling stage, and the uncertainty generated by search frictions in the labor market stage is, to the best of our knowledge, new in the literature.

These modeling choices allow us to connect the factors driving wage inequality with the endogenous choices that generate worker heterogeneity. In particular, the segmented market assumption delivers labor market sorting that manifests in two different ways. First, it generates positive assortative matching, which means that high productivity workers will typically match with high productivity firms. Second, it generates worker segregation, which means that workers with different productivity will tend to group themselves into different employers, irrespectively of the productivity of the latter (see Kremer and Maskin (1996)). These two features arise in the model because educational attainment acts as a signal of productivity. Furthermore, since the model generates a distribution of educational attainment with overlapped supports, this signal is imperfect. The connection between labor market sorting and education choices is the main novelty of this paper.

We calibrate our model using IPUMS-CPS data to match two widely used measures of wage inequality for 1980 and 2005: between-group inequality as measured by the wage premiums, and within-group inequality as measured by the variance of log wages conditional on educational attainment. The choice of the time period is motivated by the increase experienced in both measures of inequality. For instance, whereas the average college graduate made 41% than the average high school graduate in 1980, that difference increased to 108% in 2005. In turn, the variance of log wages went up for all educational groups. In the case of college graduates it went from .29 in 1980 to .44 in 2005[2]. The structure of our model allows us to quantify the contribution of labor market uncertainty and worker heterogeneity to the level of each of these measures, as well as to the change over the time period under analysis.

We find that the relative importance of labor market uncertainty and worker heterogeneity differs markedly across educational groups. For instance, labor market uncertainty is more important for the size of the wage premium between high school graduates and college dropouts than it is for the size of the premium between the former and college graduates. Furthermore, to understand the increase in the wage premiums during the period 1980-2005, the increase

[2]To give an idea of the magnitude of this increase, had the wage distribution of college graduates been log normal, it would correspond to an increase of the 90/10 income ratio from 4 in 1980 to 5.5 in 2005. The difference would be larger if one considers a wage distribution with fatter right tails.

in worker heterogeneity at the moment of labor market entry is crucial. Our results indicate that most of the increase was due to better educational sorting. With respect to within-group measures of inequality, we find that in 2005, the decomposition is roughly independent from schooling attainment: worker heterogeneity explains 40% of the observed wage dispersion, between-firm wage differentials explain an extra 30%, and within-firm wage differentials explain the remaining 30%. In spite of this, the change in the variance of log wages with respect to 1980 does differ between educational groups. For college graduates, most of this change is accounted for by worker heterogeneity, whereas for high school educated workers the role of between- and within-firm wage differentials is more important.

Overall, our findings indicate that the increase in worker heterogeneity has been more important in driving the rise in inequality for high-skilled, college educated workers, whereas labor market uncertainty has been more crucial for low-skilled, high school graduates. An important contribution of this paper is that it allows to relate the importance of worker fixed effects to the degree of educational sorting prior to labor market entry. In addition, our approach has the advantage that the frictional wage dispersion will not be invariant to policy changes. To the extent that policies aiming at tampering the increase in wage inequality could also affect the degree of frictional wage dispersion, our model provides a good benchmark for policy analysis.

The rest of this paper is organized as follows. In the remainder of this section we connect this paper to the literature. Section 2 presents the model, characterizes it and discusses its implications for the decomposition of wage premiums and wage dispersion. We calibrate the model and decompose wage differentials in Section 3. In this section we also include a comparative statics exercise to assess the response of the equilibrium to a shock that resembles skill biased technological change. In Section 4 we conclude.

Related Literature

Decomposition of wage differentials has been a general issue of interest for studying inequality. For example, Lemieux (2006) documents that most of the increase in wage inequality, especially between-group inequality, between 1973 and 2005 can be explained by increases in the return to postsecondary education. Similarly, Goldin and Katz (2007) find that the increase in the wage premiums between education levels explains about 60 to 70% of the rise in wage inequality in the United States between 1980 and 2005. Autor (2014) emphasizes the role of both the supply and the demand for skills in shaping inequality in the United States. In particular, he argues that the change in the wage premiums can be generally explained by the change in the supply of college graduates. We formulate this general equilibrium argument of skilled/unskilled labor in a structural framework.

Wage decomposition is also strongly related to analyses for the sources of inequality. Krueger and Perri (2006) argue that within-group wage inequality less likely translates into consumption inequality than between-group wage inequality. Using a structural model, Huggett et al. (2011) find that individual differences existing at age 23, especially variation in human capital, are more important than are shocks received over the life as a source of variation in lifetime earnings. However, in their analysis, these initial productivity differences, which can be interpreted as between-group inequality, are treated exogenously and hence what prior forces shape them is unclear.

There have been two main approaches to conduct a decomposition of wage differentials across workers. One approach is to estimate Mincerian earnings functions, controlling for many observable characteristics of workers that represent productivity differences. Most important examples include Abowd et al. (1999) and Abowd et al. (2002). This literature typically finds that observable worker characteristics cannot explain more than one third of the variation in wages.

The other approach is to build on dynamic optimization theory and estimate structural search models. Burdett and Mortensen (1998) consider the wage-posting equilibrium search model in which wage dispersion is naturally emerged ex post as an outcome of on-the-job search[3]. One advantage of this approach compared to the one based on reduced form estimation equations is that it does not suffer from the endogeneity problem, which typically exists in the reduced form approach (Rosenzweig and Wolpin (2000)), and hence does not bias the resulting decomposition. In addition, it enables counter-factual experiments in which effects of policy changes will be assessed. Eckstein and van den Berg (2007) provide an excellent survey of the structural approach using search theory.

In this literature, there are two papers that estimate the structural search model in the same spirit of the present paper. First, Postel-Vinay and Robin (2002) consider a search model in which firms can respond to the outside job offers received by their employees. They then decompose the wage differentials into variations due to workers' productivity heterogeneity, firms' productivity differences, and search friction. Compared to their framework, in our model workers' productivity at the labor market entry is endogenously determined through educational investments rather than a purely random shock that is independent of any change in economic conditions. Therefore, in general equilibrium, we can analyze the meaningful interaction between the workers' productivity distribution and the market environment, in particular labor market frictions.

Second, Flinn and Mullins (2015) introduce schooling decisions to an otherwise standard

[3]Wage dispersion exists among identical workers as long as the firms' productivity distribution is discrete. See Mortensen (2003) for detail.

labor search model. Like in our model, they consider education as a productivity enhancing activity before entering the labor markets. However, since their schooling investment involves no uncertainty and the workers perfectly observe their ex ante productivity types, the education attainment in their model simply shifts the upper half of the initial continuous productivity distribution upwards without reshuffling, resulting in two disconnected productivity supports. Therefore, their model is probably too stylized to analyze the interaction between search friction and workers' educational attainments and answer important questions studied in this paper.

As in our paper, many papers emphasize the importance of option value of education as a driving force of educational decision and succeed to generate college dropouts (e.g., Stange (2012), Trachter (2015), and Lee et al. (2015)). The differences between these papers and our paper are twofold: (i) our focus is on the effect of the postsecondary education on the labor market outcomes in a labor search framework, and (ii) we analyze these educational decisions in a general equilibrium context.

2. Trends in Wage Inequality

This section provides the empirical evidence of the trends in inequality in the U.S.

2.1. Data

We use the Current Population Survey (CPS), relying on IPUMS-CPS. The CPS is a nationally representative data set that covers important demographic and employment information. Our sample is white males aged from 25-55. We drop women from the sample, because the educational attainments and the labor force participation of women have changed dramatically over the past 30 years for various reasons, not only the one on which we focus in this paper. We also drop non-participants of the labor and samples with missing observations. The sample weights provided are used for computing the empirical moments.

We first define education categories. For this, we use a variable for educational attainments provided by IPUMS-CPS [4]. We define high school dropouts as those with fewer than 12 years of completed schooling or those without high school diploma; high school graduates as those having 12 years of completed schooling and not reporting no diploma; some college attendees as those with any schooling beyond 12 years and less than 4 years of college; college-plus graduates

[4] This variable, called EDUC, is constructed from two other variables, HIGRADE and EDUC99. HIGRADE, available prior to 1992, gives only the respondent's highest grade completed, whereas EDUC99, available beginning in 1992, also provides data on highest degree or diploma attained. In EDUC, the categories of HIGRADE are given the same codes as their approximate equivalents of EDUC99.

as those with 4 or more years of completed schooling. We do not use the high school dropouts in the analyses.

For wage inequality measures, we use those working full-time (40+ weeks and 35+ usual hours per week) for wages and salary in the private labor force. Self-employed are excluded. All amounts are adjusted to 2009 U.S. dollar using the chained CPI. We impute average hourly wages for each observation using reported work weeks and usual hours per week. We then drop those with the imputed hourly wage falling below half of the federal minimum wage.

We follow Autor et al. (2008) for top-coding. Specifically, prior to 1988, wage and salary incomes were collected in a single variable. After 1988, those were reported in two separate variables, corresponding to primary and secondary earnings. For each of these variables, top-coded values are simply reported at the top-code maximum, except for the primary earnings variable of 1996 or after. For that, top-coded values are assigned the mean of all top-coded earners, and so we reassign the top-coded value. We then multiply the top-coded earnings value by 1.5. After 1988, we simply sum the two earnings values to calculate total wage and salary earnings.

2.2. Trends in Wage Premium and Wage Dispersion

We focus on the standard measures of wage inequality used in the literature (e.g. Lee et al. (2015)), wage premium and wage dispersion.

Figure 1 plots the trend of wage premium measured by the ratio of the average wage of high school graduates to that of college graduates / some college attendees. We take five-year averages to mitigate the cyclical effects. For example, the 1990 value is the average of 1988-1992. This measure, also known as between-group wage inequality, is most widely used in the literature as an inequality measure. The figure shows that the wage premium of college graduates has increased over time: it was around 40% in the early 1980s and has raised to around 110% in the early 2000s. A similar trend can be observed for some college attendees, but the increase is milder and has stopped in the 1990s.

Figure 2 plots the trend of wage dispersion measured by the variance of log wages for each education group, high school graduates, some college attendees, and college graduates. We again take five-year averages. The figure shows that the wage dispersion has increased over time for all the three education categories. The shape, however, looks very different for college graduates. The variance of log wages both for high school graduates and some college attendees has steadily increased from around 0.2 to 0.3 over the past 30 years. On the other hand, it had been flat for college graduates at around 0.3, but sharply increased in the early 1990s, reaching to even around 0.45, and became flat again in the late 2000s. Overall, the variance of log wages

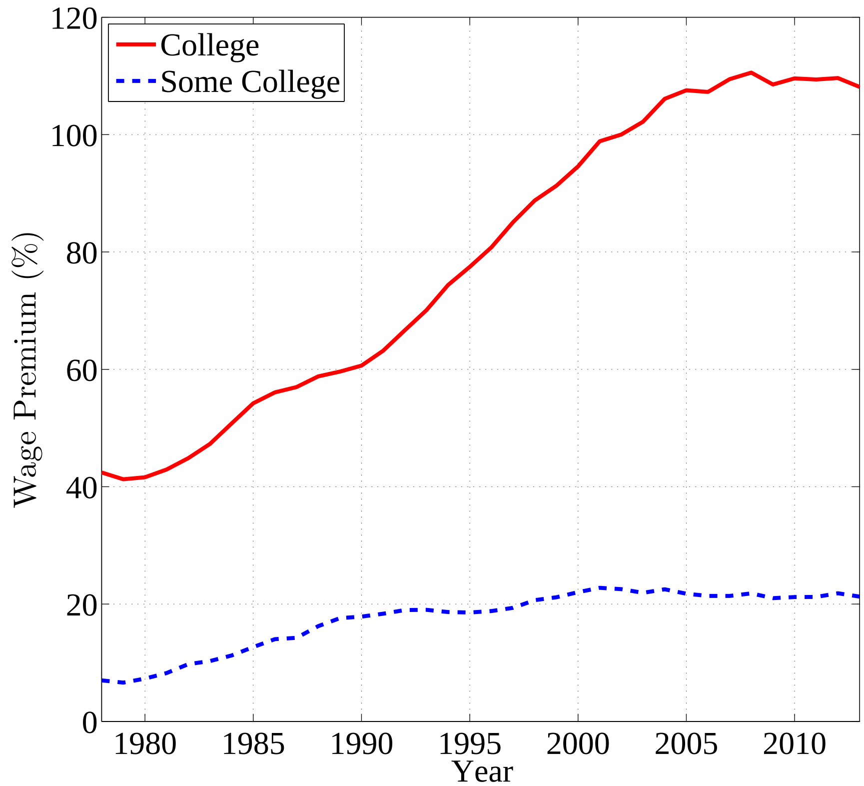

Figure 1: Wage Premium. The figure plots the wage premium of college graduates and some college attendees measured as the ratio of average wage of that education group to average wage of high school graduates (five-year average).

has become one and half times larger over the past 30 years for all the three education groups.

Lee et al. (2015) also focus on these two measures of wage inequality, and hence it is worth pointing out here the key difference between our study and theirs. They consider the wage inequality as a labor market risk, and, taking this risk *as given*, construct an education choice model to account for the pattern of supply of skills. In its spirit, on the other hand, our paper is closer to Goldin and Katz (2007) and Autor (2014) that highlight the role of supply and demand of skills in explaining the trends of wage premiums. We build on this general equilibrium argument in our structural framework, extending our focus to the wage dispersion too. Our structural model can account for the trends in supply *and* demand of skills, and hence generate the wage inequality as an equilibrium outcome, which is essential for analyzing sources of inequality and providing policy evaluations.

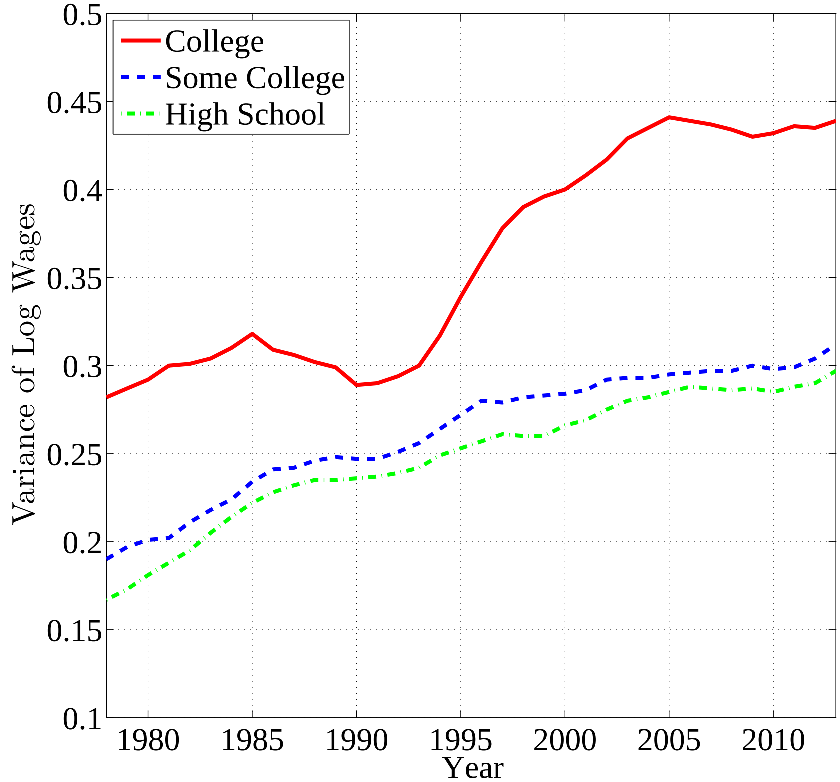

Figure 2: Wage Dispersion. The figure plots the wage dispersion of college graduates, some college attendees, and high school graduates measured as the variance of log wages (five-year average).

3. The Model

Time is continuous and infinite. There is a unit measure of workers and a measure m of firms. Both agents are risk neutral and discount the future at the common discount rate r. Workers face a constant birth/death rate μ, whereas firms live forever. Workers' lifetime is divided in two stages: a schooling stage and a labor market stage. In the schooling stage, workers enhance their human capital prior to labor market entry. In the labor market stage, workers and firms are matched randomly and production takes place.

3.1. Schooling Stage

Workers start their life with a productivity $z \in \mathcal{Z}$ and an ability $a \in \mathcal{A}$ to complete schooling levels. Schooling ability a is time-invariant, and productivity z evolves as the agent completes schooling levels. For simplicity, we restrict ourselves to $\mathcal{A} = \{a_0, a_1\}$ where $a_0 < a_1$.

Schooling levels are indexed by $i \in \mathcal{I} \equiv \{1, 2, ..., I\}$, each associated with a vector $\{\theta_i, \varphi_i, c_i\}$.

We model the completion of schooling level i as a Poisson process with arrival rate $\theta_a^i \equiv \theta_i a$. The parameter θ_i measures how long it takes to complete schooling level i, whereas the ability level a indicates how likely the worker would indeed complete it. Upon completion of schooling level i, workers' productivity increases from z to $(1+\varphi_i)z$. While no completion occurs, workers incur an instantaneous marginal utility cost c_i.

Workers know their productivity level z, but not their ability type a. Each worker holds the prior belief q that they are the high ability type, i.e. $q = \Pr[a = a_1]$, and update this prior based on Bayes' rule. To understand how the belief is updated, consider a worker with q_t at time t. The probability of her finishing the schooling level i over the next time interval dt is given by $\theta_a^i dt$. Thus, when she does not succeed, q_t becomes

$$q_{t+dt} = \frac{q_t(1-\theta_1^i dt)}{q_t(1-\theta_1^i dt) + (1-q_t)(1-\theta_0^i dt)}$$

Subtracting q_t from both sides and dividing by dt and then taking $dt \to 0$, we get

$$\frac{dq_t}{dt} = -\Delta\theta^i(1-q_t)q_t \tag{1}$$

where $\Delta\theta^i \equiv \theta_1^i - \theta_0^i$. If she succeeds, the prior jumps to

$$\hat{q}(q_t) = \frac{\theta_1^i}{\theta^i(q)} q_t \tag{2}$$

where $\theta^i(q) \equiv q_t\theta_1^i + (1-q_t)\theta_0^i$. In other words, the prior appreciates discretely if completion occurs and depreciates continuously otherwise.

Workers hold the option to dropout schooling and enter immediately the labor market. The decision to dropout is assumed to be irreversible. This means workers' problem corresponds to an optimal stopping problem in which they must decide when to dropout schooling. Their optimal decision will take the form of a cutoff belief $\overline{q}^i : \mathcal{Z} \to [0,1]$ that will trigger dropout while pursuing schooling level i if and only if $q_t < \overline{q}^i(z)$.

The individual state of the worker in the schooling stage is given by the triple (q, z, i). We use $S^i : [0,1] \times \mathcal{Z} \to \mathbb{R}$ to denote the value of pursuing schooling level i. For $i < I$, if the option to dropout has not been exercised, S^i must satisfy the following stochastic partial differential equation:

$$\rho S^i(q,z) = \rho(1-c_i)bz + S_q^i(q,z)\frac{dq}{dt} + \theta^i(q)\left[S^{i+1}(\hat{q}(q),(1+\varphi_i)z) - S^i(q,z)\right] \tag{3}$$

where the subscript q on S_q^i denotes the partial derivative of S^i with respect to q and we used

$\rho \equiv r+\mu$. The first term on the right hand side is the instantaneous utility, which we normalized by the effective discount factor ρ for convenience. The second term captures the change in the value of schooling as a consequence of the deterioration of the worker's belief. The last term represents the expected change in the value of schooling upon completion of schooling level i. In this case, she moves to the next schooling level $i+1$ with a new prior and the increased productivity. In the case $i = I$, completion means that the worker enters the labor market, which implies that S^I must satisy:

$$\rho S^I(q,z) = \rho(1-c_I)bz + S^I_q(q,z)\frac{dq}{dt} + \theta^I(q)\left[U^I((1+\varphi_I)z) - S^I(q,z)\right]. \tag{4}$$

where $U^I(z)$ denotes the value of being unemployed in submarket I for a worker with productivity z. Upon completion, the subjective belief q becomes irrelevant. The important thing to notice is that optimal schooling decisions will generate endogenously a non-degenerate distribution of workers over pairs (z,i) at the moment of labor market entry.

3.2. Labor Market Stage

Labor markets are indexed by the level of schooling attainment. In this stage, workers and firms are matched randomly and production takes place. The individual state of a worker upon entry to the labor market consists of the pair (z,i), where $z \in \mathcal{Z}$ denotes worker's productivity and $i \in \mathcal{I}$ denotes his schooling attainment. In this stage, both z and i are time invariant and perfectly observable for every agent. We use $\Psi(z,i)$ to denote the fraction of workers that participate in submarket i and have productivity level less than or equal to z. In turn, we use $\Psi^i(z)$ to denote the distribution of productivity levels z conditional on being in submarket i.

Firms are also heterogeneous in their productivity level, which is denoted by $p \in \mathcal{P} \equiv [b, \overline{p}]$. The production technology is linear in efficiency units of labor: the marginal product of type-p firm of hiring an additional worker with labor productivity z is given by $ph(z)$, where the function $h : \mathbb{R} \to \mathbb{R}$ maps productivity levels into efficiency units. The total output of a firm with productivity p is equal to p times the sum of its employees' efficient units of labor.

Wage Determination

We adopt Cahuc et al. (2006) wage bargaining framework, which we describe here succinctly and refer the reader to that paper for further details. Workers participating in submarket i sample job offers at a rate λ_i^U if unemployed and λ_i^E if employed. The type of the firm from which the offer originates is drawn from the sampling distribution F^i, which we will describe later. Upon matching, bargaining occurs in a context of complete information. We use β_i to denote workers' bargaining power and assume, without loss of generality, that wages are set in terms of efficient units of labor. Contracts can only be renegotiated under mutual agreement.

A worker matched with a type-p firm can be either employed or unemployed. If the worker is unemployed, the wage resulting from the bargaining process between the worker and the firm is denoted by $\phi_0(p)$. If the worker is employed in an incumbent firm with productivity p, the type-p' firm successfully poaches the worker if and only if $p < p'$. In this case, the resulting wage is denoted $\phi(p, p')$, where the first argument represents the productivity of the last employer and the second argument is the productivity of the new employer. In the case in which the firm with productivity p can successfully deter poaching by a type-p' firm, e.g. if $p > p'$, renegotiation results in a wage raise within the firm if an only if $p' \in (q(w, p), p]$ where $q(w, p)$ is defined as the productivity level that satisfies:

$$\phi(q(w, p), p) = w \tag{5}$$

The productivity cutoff $q(w, p)$ can be interpreted as the maximum firm productivity that does not trigger a wage raise in a match in which a type-p firm pays a wage equal to w. In the remainder of the description of the labor market stage we take these wage functions as given.

Firms

A firm that wants to hire a worker with schooling level i must create a vacancy in submarket i. The firm is then randomly matched with a worker and they bargain over the wage. A match between a worker and a firm is destroyed at a rate δ_i, which is exogenously given and specific to submarket i.

Let $\pi^i(w, p, z)$ denote the profit per worker for a firm with productivity p when hiring a worker of type (z, i) at wage w. This function can be expressed recursively as follows:

$$\begin{aligned} \rho\pi^i(w, p, z) = \quad & \rho(p - w)h(z) - (\delta_i + \lambda_i^E(1 - F^i(p)))\pi^i(w, p, z) \\ & + \lambda_i^E \int_{q(w,p)}^{p} (\pi^i(\phi(x, p), p, z) - \pi^i(w, p, z))dF^i(x) \end{aligned} \tag{6}$$

where, as we mentioned above, λ_i^E and F^i represent the employed worker's contact rate and the sampling distribution of potential employers in submarket i, respectively. The first term in the right hand side is the flow profit. The second term is the capital loss that stems from separation, or poaching, which occurs when the worker receives an offer from a firm with higher productivity than its current employer. The third term, which is negative, captures the capital loss that occurs when the incumbent employer can successfully deter poaching, but at the cost of a wage raise.

Since vacancies are posted before matching occurs, the incentives to create a new vacancy depend on expected profits rather than on realized profits. To obtain an expression for the expected profits function, notice that, upon matching, there are three possible outcomes: the vacancy is filled by an unemployed worker, the vacancy is filled by a worker previously employed at a firm with lower productivity, or the vacancy remains unfilled. Hence, the expected profit from posting a vacancy in submarket i must satisfy the following recursive expression:

$$\begin{aligned}\rho\Pi_0^i(p) = \rho\pi_0 \quad &+ \quad \frac{u_i}{\lambda_i^U}\left[\int_{z\in\mathcal{Z}}\pi^i(\phi_0(p),p,z)d\Psi^i(z) - \Pi_0^i(p)\right] \\ &+ \quad \frac{1-u_i}{\lambda_i^E}\left[\int_{z\in\mathcal{Z}}\left(\int_{x\in\mathcal{P}}\pi^i(\phi(x,p),p,z)l_i(x)dx\right)d\Psi^i(z) - \Pi_0^i(p)\right] \end{aligned} \tag{7}$$

where π_0 is the flow payoff (cost) of having a vacancy posted, u_i is the unemployment rate in submarket i and $l_i(p)$ is the mass of workers employed in a type-p firm. The second term in the right hand side represents the profit if the firm matches with an unemployed worker, which occurs with probability u_i/λ_i^U. The third term represents the profit if the firm matches with an employed worker, which occurs with probability $(1-u_i)/\lambda_i^E$.

A firm with productivity p takes as given $\{l_i, \lambda_i, F^i\}$ for all i, and the wage equations, and chooses a vacancy creation policy $v_i(p)$ in order to solve the following problem:

$$\max_{v_i}\left[\Pi_0^i(p)v_i - \kappa(v_i)\right] \tag{8}$$

for each i, where κ is the vacancy creation cost function, which is assumed to be strictly increasing and strictly convex.

Workers

Workers are either unemployed or employed. Workers with schooling attainment i search for a job in submarket i. Let $W^i(w,p,z)$ be the value of being employed at wage w in a type-p firm for a type-z worker, and $U^i(z)$ be the value of being unemployed for a type-z worker. These values must jointly satisfy the following functional equations:

$$\begin{aligned}\rho W^i(w,p,z) = \quad & \rho w h(z) + \delta_i(U_i(z) - W^i(w,p,z)) \\ & +\lambda_i^E\int_{q(w,p)}^{p}\left(W^i(\phi(x,p),p,z) - W^i(w,p,z)\right)dF^i(x) \\ & +\lambda_i^E\int_{p}^{\overline{p}}\left(W^i(\phi(p,x),x,z) - W^i(w,p,z)\right)dF^i(x)\end{aligned} \tag{9}$$

$$\rho U_i(z) = \quad \rho b(z) + \lambda_i^U\int_b^{\overline{p}}\left(W(\phi_0(x),x,z) - U_i(z)\right)dF^i(x) \tag{10}$$

The first term in the right hand side of (9) is the wage paid. The second term represents the net loss due to exogenous separation. The third term corresponds to the expected net gain from a wage increase by the current employer. The last term is the expected net gain from accepting a higher wage offer made by a different employer. Likewise, the first term in the right hand side of (10) is the flow value of leisure that is linear in z and independent of i. The second term is the expected net gain from accepting a new offer.

Sampling Distribution and Contact Rates

We assume that the probability of being matched with a type-p firm in submarket i is proportional to the number of vacancies each firm posts in each submarket. Thus, the sampling distribution function must satisfy:

$$F^i(p) = \frac{\int_b^p v_i(x)d\Gamma(x)}{\int_b^{\overline{p}} v_i(x)d\Gamma(x)} \tag{11}$$

where $v_i(p)$ is the optimal number of vacancies posted that solves (8), and Γ represents firms' exogenous productivity distribution. This equation makes precise that the sampling distribution in our model is an endogenous object.

From workers' perspective, the contact rate represents the probability she will be matched with a prospective employer in submarket i. This rate is given by:

$$\lambda_i^U = \frac{\int_b^{\overline{p}} v_i(x)d\Gamma(x)}{\int_{z\in\mathcal{Z}} d\Psi(z,i)} m \tag{12}$$

The expression in the numerator represents the number of vacancies posted by all firms participating in submarket i, whereas the denominator represents the number of workers looking for a job in the same submarket. Observe that both elements are derived from optimal decisions of economic agents. In fact, the novelty of this paper is in the denominator, since the fraction of workers participating in each submarket stems from the optimal decisions made in the schooling stage.

Free Entry

We close the model by imposing a free entry condition for firms. In order to participate in the labor market and to be able to post vacancies, a firm must pay the entry cost E. Firms are assumed to know their productivity p before entry. Therefore, a type-p firm will enter the labor market and start posting vacancies if and only if:

$$\sum_{i\in\mathcal{I}} \Pi^i(p) - E \geq 0 \tag{13}$$

where

$$\Pi^i(p) \equiv \Pi^i_0(p) v^\star_i(p) - \kappa(v^\star_i(p))$$

is the expected profit from participating in submarket i, net of vacancy creation costs, and $v^\star_i(p)$ is the solution to the vacancy creation problem (8). This condition assumes implicitly that the payment of the entry cost E is required to post vacancies in *any* submarket. Upon entry, a firm might optimally choose not to post vacancies in a given submarket. In the context of our model, this is rather unappealing since we want the measure of firms that operate in each submarket to be independent of i. A parsimonious way of achieving this goal is by making $\pi_0 = 0$, an assumption that we will maintain in the remainder of the paper.

In equilibrium, the last firm entering the market has productivity type $\tilde{p}$, which is pinned down by the following condition:

$$\sum_{i\in\mathcal{I}} \Pi^i(\tilde{p}) = E \tag{14}$$

Furthermore, as long as $\pi_0 = 0$, these firms will post vacancies in all submarkets, which implies that the measure of all participating firms is given by

$$\tilde{m} = \int_{\tilde{p}}^{\overline{p}} d\Gamma(p) \tag{15}$$

Therefore, positive entry costs ($E > 0$) imply that there is a positive mass of low productive firms that do not enter the labor market, e.g. $\tilde{m} < m$.

3.3. Stationary Equilibrium

Let $\Psi^0_a : \mathcal{Z} \to [0,1]$ denote the initial cumulative distribution function, conditional on ability level a and ψ^0_a be the corresponding density function. The initial prior belief, denoted $q_0(z)$, corresponds to the subjective probability that each worker attaches to being of type a_1, which is calculated using Bayes' Rule as follows:

$$q_0(z) = \frac{\psi^0_{a_1}(z)}{\Pr(a = a_1)\psi^0_{a_1}(z) + (1 - \Pr(a = a_1))\psi^0_{a_0}(z)} \tag{16}$$

Therefore, the individual state of a worker that starts the schooling stage is given by $(q_0, z_0, 0)$, where z_0 is drawn from Ψ_a^0, and q_0 is calculated according to (16).

A *Stationary Equilibrium* for this economy, takes as given the productivity distribution Ψ_a^0, and consists of wage functions $\{\phi_0, \phi\}$, value functions for the worker $\{S^i, W^i, U^i\}$ and the firm $\{\pi^i\}$, policy functions for the worker $\{\overline{q}^i\}$ and the firm $\{v_i\}$, equilibrium contact rates $\{\lambda_i^U, \lambda_i^E\}$, workers' distribution Ψ and a cutoff for firms' productivity $\tilde{p}$, such that:

1. the wage equations solve the wage bargaining problem;
2. the policy functions solve workers' schooling decision problem and firms' vacancy creation problem;
3. the value functions satisfy their respective recursive expression;
4. the equilibrium contact rates are consistent with individual choices;
5. workers' distribution is stationary; and
6. the productivity cutoff $\tilde{p}$ satisfies the free entry condition.

In the next section we establish the linearity of the stationary equilibrium, perform a partial characterization of it, and discuss the implications of these results for wage premium and wage dispersion.

4. Equilibrium Characterization

Unemployment Rates

Worker flows endogenously determine the equilibrium unemployment rate, u^i. In a stationary environment, the ins and outs of unemployment must be equal to each other. The ins are given by $(1 - u^i)\delta_i + \mu$, which represents workers that lose their jobs due to job destruction and that either complete or abandon schooling. In turn, the outs are $\left(\lambda_i^U + \mu\right) u^i$, which represents workers that either find a job or die. The stationary condition for this reads

$$u^i = \frac{\delta_i + \mu}{\delta_i + \mu + \lambda_i^U}. \tag{17}$$

The unemployment rate is increasing in the job destruction rate and decreasing in the job finding rate.

Wage Equations

We borrow from the results on Cahuc et al. (2006). Consider a firm with productivity p. If the contacted worker is employed in a firm with productivity $p' > p$, the incumbent employer can deter poaching and any wage offer will be rejected by the worker. We assume then that $\phi(p', p) = p$. If the contacted worker is employed in a firm with productivity $p' < p$, the solution to the bargaining problem is given by

$$\phi(p', p) = p - (1 - \beta_i) \int_{p'}^{p} \frac{\rho + \delta_i + \lambda_i^E(1 - F^i(x))}{\rho + \delta_i + \lambda_i^E \beta_i(1 - F^i(x))} dx \tag{18}$$

Finally, if the contacted worker is unemployed, the solution to the bargaining problem is given by:

$$\phi_0(p) = p - (1 - \beta_i) \int_{b}^{p} \frac{\rho + \delta_i + \lambda_i^E(1 - F^i(x))}{\rho + \delta_i + \lambda_i^E \beta_i(1 - F^i(x))} dx \tag{19}$$

We will use these equations to characterize the profit per worker.

Profit per Worker, Optimal Vacancies and Free Entry

In this section we show that that the profit per worker is linear in worker's productivity z and use this result to characterize the optimal vacancy creation policy. We start with the following proposition.

Proposition 1 *The profit per worker is linear in worker's productivity, i.e.* $\pi^i(w, p, z) = \pi^i(w, p)h(z)$ *where:*

$$\pi^i(w, p) = (1 - \beta_i) \int_{q(w,p)}^{p} \frac{\rho}{\rho + \delta_i + \lambda_i^E \beta_i(1 - F^i(x))} dx \tag{20}$$

All proofs are relegated to the appendix. It follows that we can write the value of an unfilled vacancy posted in submarket i as follows:

$$\Pi_0^i(p) = \frac{1}{\rho + \frac{u_i}{\lambda_i^U} + \frac{1-u_i}{\lambda_i^E}} \left(\rho\pi_0 + \left[\frac{u_i}{\lambda_i^U} \pi^i(\phi_0(p), p) + \frac{1-u_i}{\lambda_i^E} \int_{x \in \mathcal{P}} \pi^i(\phi(x, p), p) l_i(x) dx \right] \mathrm{E}[h(z)] \right) \tag{21}$$

where the expectation is taken with respect to workers' productivity distribution in submarket i. Thus, the vacancy creation problem becomes:

$$\max_{v_i} \left[\Pi_0^i(p)v_i - \kappa(v_i)\right] \tag{22}$$

for all i. If we asume further that the vacancy creation cost is iso-elastic:

$$\kappa(v) = \chi \frac{v^{1+\frac{1}{\xi}}}{1+\frac{1}{\xi}}$$

we obtain the following expression for the number of vacancies created by a type employer of type p in submarket i:

$$v_i^\star(p) = \left(\frac{\Pi_0^i(p)}{\chi}\right)^\xi \tag{23}$$

Given the expression for $\Pi_0^i(p)$ in (21), this equation shows that the vacancy creation policy in submarket i does not depend on the entire distribution of worker's productivity within it, but just on its conditional mean. This property of the optimal policy will prove useful in the computation of equilibrium. Finally, notice that

$$\Pi^i(p) \equiv \Pi_0^i(p)v_i^\star(p) - \kappa(v_i^\star(p)) = \frac{\Pi_0^i(p)^{1+\xi}}{(1+\xi)\chi^\xi} \tag{24}$$

Hence, the free entry condition can be written as

$$\sum_{i\in\mathcal{I}} \frac{\Pi_0^i(\tilde{p})^{1+\xi}}{(1+\xi)\chi^\xi} = E \tag{25}$$

Notice that if there is no cost of keeping the vacancy posted, e.g. $\pi_0 = 0$, then any firm that participates in the labor market will post vacancies in *all* submarkets.

Value of being unemployed

The following proposition shows that the value of being unemployed is linear in worker's productivity z and it depends on the contact rates that prevail in each submarket. The mechanism studied in this paper hinges on the connection between these two equilibrium objects, e.g. the value of unemployment and the contact rates.

Proposition 2 *The value of being unemployed is linear in worker's productivity, i.e.* $U^i(z) =$

$U^i z$, *where:*

$$U^i = b + \beta_i \int_b^{\overline{p}} \frac{\lambda_i^U (1 - F^i(x))}{\rho + \delta_i + \lambda_i^E \beta_i (1 - F^i(x))} dx \tag{26}$$

Observe that when workers have no bargaining power, e.g. $\beta_i = 0$, the value of being unemployed is independent of the contact rate. Moreover, the sensitivity of the value of being unemployed with respect to the contact rate while unemployed λ_i^U is increasing in the bargaining power of workers. In the context of our model, this implies the unemployment value in submarket i is more sensitive to worker flows and consequently, schooling decisions are also more sensitive to participation in that submarket.

It is also important to notice that if workers lack of any bargaining power, there would be no connection between unemployment values and contact rates. This case would correspond to Postel-Vinay and Robin (2002) and it is a consequence of both the Bertrand competition assumption and the full information assumption (e.g. employers know everything about the workers they contact). Employers are able to appropriate the option value of searching in a frictional labor market. This feature of the environment stands in contrast with Burdett and Mortensen (1998) paradigm, in which wage offers are made behind the veil of ignorance, e.g. employers do not know the type of the worker they will contact. Although this assumption restores the connection between unemployment values and contact rates, it fails to deliver wage dispersion within the firm, which, as we will see, corresponds to the type of dispersion labeled as frictional.

Optimal policies in the Schooling Stage

We can use (3) and (4) to characterize the cutoff $\overline{q}^i$ that triggers dropout. At this cutoff value, the worker must be indifferent between continuing education or entering labor market. In addition, optimality requires the marginal benefit of continuing education for one more instant to be equal to the marginal cost of delaying labor market entry. These two requirements are represented by the following value matching and smooth pasting conditions:

$$S^i(\overline{q}^i(z), z) = U^{i-1} z, \tag{27}$$

$$S_q^i(\overline{q}^i(z), z) = 0, \tag{28}$$

for all $i \in \mathcal{I}$. The make further progress, we will use the following lemma, which states that value of schooling can be expressed as a linear function of z.

Lemma 1 *The value of schooling can be written as* $S^i(q, z) = S^i(q) z$ *for all* i*, where* $S^i(q)$

satisfy the corresponding versions of (3) and (4):

$$\rho S^i(q) = \rho(1-c_i)b - \Delta\theta^i(1-q)qS^i_q(q) + \theta^i(q)\left[(1+\varphi_{i+1})S^{i+1}(\hat{q}(q)) - S^i(q)\right], \tag{29}$$

$$\rho S^I(q) = \rho(1-c_I)b - \Delta\theta^I(1-q)qS^I_q(q) + \theta^I(q)\left[(1+\varphi_I)U^I - S^I(q)\right]. \tag{30}$$

The proof is straightforward and only requires to plug $S^i(q)z$ into (3) and (4). The next proposition uses this result to characterize the optimal cutoff beliefs.

Proposition 3 *A worker pursuing schooling level $i \in \mathcal{I}$ drops out education whenever her belief regarding her own ability falls below $\overline{q}^i$ which must satisfy:*

$$\overline{q}^i = \frac{\rho(U^{i-1} - (1-c_i)b) - \theta_0\left[(1+\varphi_i)S^{i+1}(\hat{q}(\overline{q}^i)) - U^{i-1}\right]}{\Delta\theta^i\left[(1+\varphi_i)S^{i+1}(\hat{q}(\overline{q}^i)) - U^{i-1}\right]}, \tag{31}$$

$$\overline{q}^I = \frac{\rho(U^{I-1} - (1-c_I)b) - \theta_0\left[(1+\varphi_I)U^I - U^{I-1}\right]}{\Delta\theta^I\left[(1+\varphi_I)U^I - U^{I-1}\right]}. \tag{32}$$

This result is very useful because it clarifies that cutoff rules do not depend on workers' productivity. This implies that two workers with the same initial productivity can end up with different educational attainment, which is a consequence of the informational structure of the model.

Accepted Wage Distribution within the Firm

To characterize the accepted wage distribution we follow Cahuc et al. (2006) closely. In this section we only display the relevant expressions and refer the reader to that paper for details. Let $G^i(w \mid p)$ denote the mass of workers employed in a type-p firm that earn a wage less or equal than w. In a stationary equilibrium, this mass if given by:

$$G^i(w \mid p) = \left(\frac{\delta_i + \mu + \lambda_i^E(1 - F^i(p))}{\delta_i + \mu + \lambda_i^E(1 - F^i(q(w,p)))}\right)^2 \tag{33}$$

Additionally, we use $L(p)$ to denote the fraction of workers employed in a firm with productivity less or equal than p and use $l(p)$ to denote the corresponding density (e.g. the fraction of workers employed in a firm with productivity p). These functions are given by

$$L^i(p) = \frac{F^i(p)}{\delta_i + \mu + \lambda_i^E(1 - F^i(p))} \tag{34}$$

$$l^i(p) = \frac{\delta_i + \mu + \lambda_i^E}{\left(\delta_i + \mu + \lambda_i^E(1 - F^i(p))\right)^2} f(p) \tag{35}$$

Using these functions, the unconditional distribution of accepted wages is given by

$$\Upsilon^i(w) = \int_b^{\overline{p}} G^i(w \mid p) dL^i(p) \tag{36}$$

which corresponds to the mass of workers that earn a wage less or equal than w, regardless of the firm in which they work in.

5. Implications for Wages and Measures of Inequality

Let $w(z, p, i)$ be the average wage paid by a firm with productivity p to a worker of type z and schooling attainment i. The model laid out in the previous section implies that:

$$w(z, p, i) = \mathrm{E}[\phi_i(p', p) \mid p, i] h(z, i) \tag{37}$$

where p' is the productivity of the best alternative for the the worker, and the expectation is taken with respect to the distribution of workers within the firm. Recall that ϕ is the wage per efficient units of labor and h is worker's productivity. The average wage paid in submarket i is given by:

$$\mathrm{E}[w(z, p, i) \mid i] = \mathrm{E}[\omega(p, i) h(z, i) \mid i]$$

where $\omega(p, i) \equiv \mathrm{E}[\phi_i(p', p) \mid p, i]$. Conditioning on an educational level implies that h and ω are uncorrelated, which allows us to write:

$$\mathrm{E}[w(z, p, i) \mid i] = \mathrm{E}[\omega(p, i) \mid i] \mathrm{E}[h(z, i) \mid i]$$

This equation highlights the approach taken in this paper. Most macroeconomic models studying wage inequality assume that the first term in the right hand side is exogenously given and focus on the process of human capital accumulation that shapes the second term. In turn, search models of the labor market focus on the first term, assuming that worker heterogeneity at the moment of labor market entry is exogenous. The key feature of our model is to consider both elements as being endogenously determined.

In the quantitative section, we shall assume that $h(z) = \exp(z)$. Using this functional form in the previous equation delivers:

$$\mathrm{E}[w(z, p, i) \mid i] = \mathrm{E}[\omega(p, i) \mid i] \times h(R_i) \times \mathrm{E}[h(z_0, i) \mid i] \tag{38}$$

where we have introduced the notation $R_i = \prod_{j=0}^{i}(1 + \varphi_j)$. The average wage in submarket i

is the product of three terms. The first term is related to labor market conditions.The second term is related to human capital accumulation and it captures how labor market compensation increases as a result of a higher educational attainment. The third term corresponds to the average efficiency units offered by workers, conditional on educational attainment. Notice that the conditional expectation is meaningful to the extent that the probability of completing college education is correlated with the initial productivity z_0. It is important to point out how these terms are related. In our model, education not only enhances workers' productivity but also sorts workers into different submarkets according to their productivity. By doing so, it influences labor market conditions through its effect on vacancy creation in each submarket (see Equations (21) and (23)). In addition, labor market conditions also affect educational choices because they determine the value of unemployment in each submarket (see Equation 10).

Equation (38) allows us to relate the size of the wage premiums to each of the three components mentioned in the previous paragraph. In our model, wage premiums (relative to the lower schooling level $i = 0$) are given by:

$$1 + \text{premium}_i = \frac{\mathrm{E}[\omega(p, i) \mid i]}{\mathrm{E}[\omega(p, 0) \mid 0]} \times \frac{h(R_i)}{h(R_0)} \times \frac{\mathrm{E}[h(z_0, i) \mid i]}{\mathrm{E}[h(z_0, 0) \mid 0]} \tag{39}$$

for any level of educational attainment $i \in \mathcal{I}$. This expression allows us to identify the contribution of worker's heterogeneity and labor market uncertainty to the magnitude of the wage premiums. Workers' heterogeneity at the moment of labor market entry is shaped by the choices made at the schooling stage. These decisions affect the size of these premiums through two different channels: by enhancing worker's productivity and by sorting workers according to their innate productivity into different educational groups. Productivity enhancement is measured by φ_i and it is exogenous in the model. The degree of sorting is determined endogenously by the learning process, the heterogeneity in z_0 and the pattern of correlation between productivity and ability[5].

Notice that according to Equation (39), if we assume no heterogeneity in z_0, the contribution of worker's productivity to the magnitude of the wage premiums would be fully determined by the returns to education. Productivity differences being dampened in the labor market is a theoretical possibility. In fact, labor market uncertainty exacerbates productivity differences if and only if the first term in (39) is larger than one. As we will see later, our model suggests that as long as there is educational sorting, we should expect productivity heterogeneity to be exacerbated by the labor market.

To analyze the implications of our model for the variance of log wages, we write equation

[5] The role of education in shaping worker's productivity has been modeled in two alternative ways. Lee et al. (2015) and the references therein see educational technology as purely productivity revealing. Other papers see it as a productivity enhancing technology. Our model encompasses both views.

(37) in log terms as follows:

$$\hat{w}(z,p,i) = \hat{h}(z,i) + \hat{\omega}(p,i) \tag{40}$$

where we are using the notation $\hat{x} = \ln x$. This equation is related to the typical wage equation estimated using the methodology laid out in Abowd et al. (1999). The two terms in the right hand side can be interpreted as the firm and worker fixed effects, respectively. Using the law of total variance, we can decompose the unconditional variance of log wages as follows[6]:

$$\begin{aligned} \mathrm{Var}\left[\hat{w}(z,p,i)\right] &= \mathrm{Var}[\mathrm{E}[\hat{h}(z,i) \mid p]] + \mathrm{Var}[\mathrm{E}[\hat{\omega}(p,i) \mid p]] \\ &+ 2\mathrm{Cov}[\hat{h}(z,i), \hat{\omega}(p,i)] + \mathrm{E}[\mathrm{Var}[\hat{h}(z,i) \mid p]] + \mathrm{E}[\mathrm{Var}[\hat{\omega}(p,i) \mid p]] \end{aligned} \tag{41}$$

The first term in the right hand side is often interpreted as an indicator of worker segregation: the fact that workers with different skills group into different firms. Notice that the only source of variation in this term comes from the fact that the expected h is different across submarkets. The second term measures wage inequality induced by between-firm dispersion. The covariance between firm and worker fixed effects is an indicator of the degree of assortative matching induced by the model. Finally, the last two terms capture the dispersion of wages within a firm which stems from both worker heterogeneity and labor market frictions. The last term would correspond to the residual wage term in wage equations. In the model, similar workers working in similar firms could end up earning different wages because they have different employment history. This can be seen more clearly by noticing that we are not conditioning wages by the productivity of worker's last second best option.

Notice that since the average wage per efficient units of labor ω does not depend on productivity, and the number of efficiency units do not depend on firm's productivity, $\hat{h}$ and $\hat{\omega}$ are uncorrelated when we condition on schooling attainment. Hence, the variance of log wages conditional on educational attainment is given by:

$$\mathrm{Var}[\hat{w}(z,p,i) \mid i] = \quad \mathrm{Var}[\hat{h}(z,i) \mid i] + \mathrm{Var}[\mathrm{E}[\hat{\omega}(p,i) \mid p] \mid i] + \mathrm{E}[\mathrm{Var}[\hat{\omega}(p,i) \mid p] \mid i] \tag{42}$$

where the covariance term dropped because of the lack of correlation between h and ω. Taken together, equations (41) and (42) highlight two important features of the model. First, to the extent that h is fully determined in the schooling stage, educational sorting is fully responsible for the degree of worker segregation displayed by the model. Second, all assortative matching in the model occurs as a consequence of the segmented market assumption[7]. Focusing on (42),

[6] The law of total variance states that $\mathrm{Var}[x] = \mathrm{Var}[\mathrm{E}[x \mid y]] + \mathrm{E}[\mathrm{Var}[x \mid y]]$

[7] Assortative matching can also be obtained within submarkets by considering more general functional forms for the production function. See for instance Lopes de Melo (2015).

we follow Postel-Vinay and Robin (2002) and label the first term the person effect, the second term the firm effect since it measures the dispersion of wages across firms, and the last term the friction effect since it measures the dispersion of wages within the firm.

6. Calibration

This section describes our calibration strategy. We calibrate the model to the U.S. economy. We consider the economy is at stationary equilibria in 1980 and in 2005, as in Lee et al. (2015). Broadly speaking, the former is interpreted as the benchmark pre-SBTC situation, whereas the latter can be seen as a post-SBTC comparison. For details of our numerical implementation, see Appendix. The data used for empirical moments are described in Section 2.

6.1. Pre-determined Parameters

We approximate our continuous-time economy with discrete time assuming that one time interval corresponds to one month.

Several parameters are identified without using the model and assumed fixed over time. The interest rate r is set so that the annual rate is 6%. The birth/death rate μ is chosen to reflect a working life expectancy of 40 years. These imply the annual discount rate of 0.92. The value of leisure b is normalized to 1.

We consider three schooling stages and thus three submarkets, $i = \{0, 1, 2\}$. Workers are born as high school students. If they dropout without breakthroughs, they are high school graduates, i.e., $i = 0$. If they dropout after one breakthrough, they become college dropouts (or some college), i.e., $i = 1$. Workers obtaining two breakthroughs become college graduates, i.e., $i = 2$. This classification is only for the sake of interpretation and hence innocuous, since the productivity does not change over time unless one gets a breakthrough.

Workers forgo a half of the flow utility while in the school, $c_1 = c_2 = 0.5$. This marginal cost captures various unmodeled costs associated with education, such as financial cost, opportunity cost, and psychic cost. We conducted a sensitivity for this parameter and found that the results are materially identical.

The common frequency of graduation (θ_1, θ_2) is set so that the average duration of schooling in the model equals 5 years. This is consistent with the length of the college education in the U.S. For example, among 2007-08 first-time bachelor's degree recipients, the median number of months between initial postsecondary enrollment and bachelor's degree attainment is 52 months [8] We assume that workers spend 1.5 years to become a college dropout, and 3.5 years to finish

[8]National Center for Education Statistics, "Profile of 2007-08 First-Time Bachelor's Degree Recipients in

Table 1: **Pre-determined Parameters Values**

Parameter	**Symbol**	**Value/Target**
Discount factor	r	Annual interest rate = 6%
Birth/Death rate	μ	Expected working periods = 40 yrs
Value of Leisure	b	1
# of Schooling Levels	I	2
Arrival rates of breakthroughs	(θ_1, θ_2)	School duration = 5 yrs (1.5+3.5 yrs)
Learning ability	(a_0, a_1)	$(0.8, 1.2)$
Fraction of high type	$\Pr(a = a_1)$	0.5
Marginal cost of schooling	c	0.5

the college, on average, reflecting the fact that a number of students leave the college without earning a certificate.

We assume the high (low) types graduate 20% faster (slower) than the average, $(a_0, a_1) = (0.8, 1.2)$. This means that, conditional on no dropping out, the high types graduate the college in 4 years on average, whereas the low types spend 6 years on average. There are equally many high types and low types in the economy. The baseline parameterization is summarized in Table 1.

6.2. Calibrated Parameters

The joint calibration of the remaining parameters takes two sub-steps in order to reduce the computational burden. A subset of parameters are exactly identified by targeting the corresponding empirical moments. We then calibrate the rest of the parameters by minimizing the RMSE.

In the following, we explain in detail how to choose the empirical moments that identify the corresponding model parameters. The targeted moments and the model moments are summarized in Table 2 and Table 3.

6.2.1. Returns to Education

Workers' educational attainment is mainly determined by the gains from the education, "returns to education". To see this, the expected college wage for a worker of type z_0 is simply given by $\mathrm{E}(w_i|i = 2)h(z_0)\prod(1 + \varphi_i)$, and absent from any change in the labor markets conditions (the first term), the return to education (the third term) determines the education decision. Thus, the workers' educational attainment distribution is informative for the marginal gains in productivity φ_i.

2009", Table 2.8.

Table 2: **Calibrated Parameters**

Parameter	Symbol	Value		Target
		1980	**2005**	
Productivity gain	φ_1 $\varphi_1\varphi_2$	1.00 1.21	1.00 1.21	Educational attainment
Separation hazard	δ_i	.015 0.010 0.003	0.011 0.008 0.003	Unemployment rate
Bargaining power	β	.23	.23	Aggregate labor share
Vacancy creation cost	χ	.006	.002	Unemployment duration
Normal variance	σ_N^2	.095	.39	See Table 3
Exponential parameter	α_a	3.92 3.92	4.70 3.33	See Table 3
Firm's Pareto parameter	γ	18.6	12.4	See Table 3

We obtain the estimates of $(\varphi_1, \varphi_1\varphi_2)$ =(-.02,.155) for 1980 and $(\varphi_1, \varphi_1\varphi_2)$ =(.025,.154) for 2005. The model estimates *genuine* returns to education. Students would gain little by spending just a little over a year in college to become dropouts, whereas their productivity certainly increases by 15% by finishing the college and earning a degree. This is qualitatively and quantitatively in line with the micro estimates.

The genuine returns to education are substantially lower than the observed wage premiums between high school graduates and college graduates. This suggests that the productivity enhancement effect of the education is limited, and the wage premiums are mainly driven by the sorting, as it will be seen in the next section.

6.2.2. Bargaining Power and Labor Market Parameters

The expected unemployment duration in submarket i can be calculated by

$$\int_0^\infty \mu \left(\int_0^t \lambda_i \tau e^{-\lambda_i \tau} d\tau \right) e^{\mu t} dt = \frac{\lambda_i}{(\lambda_i + \mu)^2} \tag{43}$$

Given the birth/death rate μ, the unemployment duration for each education group identifies the contact rate λ_i.[9] Because the contact rate is solved by equation (12) and its denominator is also given by the empirical moment, the unemployment duration is thus informative for the determinants of the firms' vacancy creation, v^i. The vacancy creation decision is driven by the bargaining power, β, and the vacancy creation parameter, χ, as indicated in eqs. (20) and (23). Thus we calibrate β and χ so that the model replicates both (i) the expected unemployment duration and (ii) the aggregate labor share in the data.[10] Flinn and Mullins (2015) also use the aggregate labor share to pin down the bargaining parameter. Our estimate indicates β =.23. Unemployment rates for each education group identify the separation hazard rate, δ_i, through Equation 17. This gives $(\delta_0, \delta_1, \delta_2)$ of $(.015, .010, .003)$ for 1980 and $(.011, .008, .003)$ for 2005, expressed in monthly frequency terms. The separation hazard is decreasing in the level of the education, so are the unemployment rates. We assume for simplicity there is no entry cost for the labor market, $E = 0$. In our future work, we will calibrate this cost parameter and endogenize the size of the firm in each point in time.

6.2.3. Production Technology and Capital-Skill Complementarity

We consider a production technology that allows us to experiment with the SBTC. At the basic level, the SBTC can be interpreted as a technological change biased towards the skilled (Violante (2008)). One driving force of the SBTC that is common in the literature is that the production technology admits capital-skill complementarity: skilled labor is more complement with equipment capital than unskilled labor. Formally, this can be expressed by the fact that the elasticity of substitution between unskilled and equipment capital is higher than that between skilled and equipment capital (Krusell et al. (2000)).

We assume a production function nonlinear in skill $h(z) = \exp(z)$. With this specification, we can easily derive the elasticity of substitution between p and z to be $1+\frac{1}{z}$, which is decreasing in the skill z. Note that the firms' technology p can be also interpreted as the size of the capital in this context. Therefore, this specification precisely admits the capital-skill complementarity.

The linearity assumption, that is rather standard in the labor search literature, is crucial for most of the analytical results we derived in quantitative section. Our production function does not, however, exhibit the decreasing marginal returns in the labor inputs that is common in the aggregate production function approach such as Krusell et al. (2000). This is not ideal,

[9] Since Equation 43 is a quadratic equation, there are two solutions for λ for any value of unemployment duration. We choose the highest root since, as μ tends to zero, it tends to the solution with $\mu = 0$.

[10] Alternatively, we could calibrate schooling-level specific bargaining power parameters, β_i, but do not opt for it for two reasons. First, the unemployment duration is quite similar across educational groups. For example, in CPS, the duration for high school graduates versus college graduates is 16 and 15 weeks, respectively, in 1980, and 20 and 21 weeks in 2005. Second, it is more natural to think that bargaining power is schooling independent, because the firms in the model only care about workers labor productivity, z, after a match.

although the drawback is certainly limited. The decreasing marginal returns would mitigate the SBTC effect in this environment as it reduces returns to skill.

6.2.4. Technology and Productivity Distribution

We make the following parametric assumptions on the distributions. First, the firms' productivity distribution is Pareto, $\Gamma(p) = 1 - p^{-\gamma}$ with the Pareto parameter γ.

The second parametric assumption is that the initial skill distribution conditional on the learning ability is Exponentially modified Gaussian (EMG). In particular,

$$z_0|a = z^N + z^E, \text{ where } z^N \sim N(\mu_N, \sigma_N^2) \text{and } z^E \sim \exp(\alpha_a).$$

This specification has several advantages. First, this implies that the initial productivity distribution conditional on the learning ability is Pareto log-normal, $\exp(z_0)|a \sim PLN(\mu_N, \sigma_N^2, \alpha_a)$, which is widely used for estimates of the labor productivity distribution. Second, we do not need to assume that the distributions for a_0 and a_1 are systematically different; rather, we assume that the two distributions differ only in the exponential parameter, α_a. This naturally introduces a correlation between the initial skill z_0 and the learning ability a. If $\alpha_{a_1} < \alpha_{a_0}$, as is the case in our estimates, then the distribution $z_0|a_1$ tends to have a thicker right tail than the distribution $z_0|a_0$.

We set μ_N so that unconditional expected initial productivity is unity, $\mathrm{E}h(z_0) = 1$. We calibrate the remaining parameters, $\theta \equiv (\gamma, \sigma_N^2, \alpha_{a_0}, \alpha_{a_1})$. We choose θ to minimize the RMSE:

$$\min_{\theta \in \Theta} \left[M(\theta) - M_{\text{data}}\right]^{\mathrm{T}} \Omega \left[M(\theta) - M_{\text{data}}\right],$$

where M_{data} is the empirical moments and $M(\theta)$ is the corresponding model moments given the set of parameters θ. We use the identity weighing matrix Ω.

We allow θ to be time dependent, and estimate it for both 1980 and 2005. For the empirical moments and the corresponding model moments, we use the wage inequality measures. Specifically, we use the wage premium between high school graduates and some college, that between high school graduates and college graduates, and the wage dispersion for each schooling level. For the model performance, Table 3.

7. Results

The predictions of the model regarding the wage premiums and wage dispersion by educational level are summarized in Table 3. The model replicates the observed wage premiums and the

Table 3: **Model Performance**

Targeted Moments	**Data**		**Model**	
	1980	**2005**	**1980**	**2005**
Premium: SC vs. HS	7.3 %	21.8 %	9.7 %	22.4 %
Premium: CL vs. HS	41.6 %	107.5 %	40.1 %	105.8 %
Dispersion: HS	0.18	0.28	0.19	0.27
Dispersion: SC	0.20	0.29	0.22	0.31
Dispersion: CL	0.29	0.44	0.31	0.44

wage dispersion in the data very well. We use this calibration to decompose the sources of wage inequality. We start by focusing on between-group wage differentials. Table 4 shows the contribution of workers' heterogeneity and labor market uncertainty to the wage premiums by educational attainment, as measured in equation (39). These results indicate that in 1980, the wage between high school graduates and college dropouts was largely due to the presence of labor market uncertainty. Relative to that benchmark, the 2005 witnessed an increase in worker heterogeneity that introduced a larger wedge between the returns of these two educational groups. In fact, in 2005 frictions could only explain about 40% of the premium. Something similar happened to the college wage premium. In 1980, almost 40% of the premium was accounted for by returns to education, a number that aligns well with typical estimates of Mincerian regressions. Importantly, the model attributes the rest to frictions, rather than to unobserved heterogeneity. This suggests that the significance of worker fixed effects in wage equations must have increased over time. In fact, unobserved heterogeneity in 2005 could explain almost half of the observed college wage premium.

In order to account for the contribution of each term, we consider equation (39) in log terms.

The importance of frictions in the determination of wage premiums is indicative of the

Table 4: **Decomposition of Wage Premiums**

	High School vs. Some College	High School vs. College
1980	9.7	40.1
Heterogeneity	-0.2 [-1.8 %]	15.5 [38.7 %]
Return to Education	-0.2 [-1.8 %]	15.5 [38.7 %]
Sorting	-0.0 [-0.0 %]	-0.0 [-0.0 %]
Friction	9.8 [102.0 %]	21.3 [53.1 %]
2005	22.4	105.8
Heterogeneity	12.7 [56.6 %]	65.1 [61.5 %]
Return to Education	2.5 [11.0 %]	15.4 [14.5 %]
Sorting	10.2[45.6 %]	49.7[46.9 %]
Friction	8.6 [38.5 %]	24.7 [23.3 %]

Table 5: **Decomposition of Wage Dispersion by Educational Attainment**

	HS	SC	CL
1980	0.195	0.220	0.309
Person	0.098 [50.5 %]	0.098 [44.8 %]	0.098 [31.8 %]
Firm	0.042 [21.6 %]	0.051 [23.2 %]	0.058 [18.8 %]
Friction	0.054 [27.9 %]	0.070 [32.0 %]	0.153 [49.4 %]
2005	0.271	0.307	0.441
Person	0.103 [38.1 %]	0.117 [38.1 %]	0.192 [43.6 %]
Firm	0.088 [32.5 %]	0.099 [32.2 %]	0.114 [25.8 %]
Friction	0.080 [29.4 %]	0.091 [29.7 %]	0.135 [30.6 %]

presence of positive assortative matching in the labor market. It is well known that models with segmented markets deliver this feature in equilibrium. There is no positive assortative matching within a submarket because from the perspective of the worker, the probability of contacting a high productivity firm does not depend on her type (e.g. in each submarket contact rates are independent of worker's type). However, if one compares across submarkets, two workers with the same productivity level but with different educational attainment will face different probabilities of being matched with a firm with a given productivity level. This occurs because education signals productivity and incentivizes high productivity firms to post more vacancies in the market for skilled workers, e.g. college graduates.

We now turn to the implications of our model for the decomposition of within-group inequality, which is performed following Equation 42. The results are displayed in Table 5. First, in 1980 the contribution of frictions to within-group inequality increased with educational attainment. Whereas for high school graduates and college dropouts, this factor explained about 30% of the observed wage dispersion, in the case of college graduates it accounted for 50% of it. This pattern changed in 2005, where frictions explain roughly the same fraction of overall dispersion for all educational groups, about 30%. This also indicates that the change in this measure of inequality came about from different sources for each educational group. Low skilled workers (e.g. High School and Dropouts) face more dispersion as a consequence of between- and within-firm wage differentials, whereas high skilled workers (e.g. College graduates) do so because of the substantial increase in worker heterogeneity. As it was the case for the decomposition of the wage premiums, this finding points out to educational sorting prior to labor market entry as an important force in shaping the increase in wage inequality over the last 30 years.

It is important to make a final remark. The correlation between productivity and ability turns out to be crucial in the analysis. To the extent that this correlation is positive, there is also a sorting effect by which high productivity workers are more likely to finish college. This sorting mechanism increases the contribution of worker's productivity to the observed measures of wage inequality. Our results therefore indicate that this correlation has strengthened over time. Whether this phenomenon came about because of changes in the educational process or because of changes in the primitive distribution of initial productivities is something our model remains silent about.

8. Conclusions

In this paper, we develop a model of schooling investment and labor market search that generates both a distribution of skills and residual wages as equilibrium objects and use it to provide a novel decomposition of wage inequality. The key interaction on which we have focused is that between distribution of educational attainment and labor market frictions.

We now highlight two lessons from our analysis that should be useful for understanding the sources of inequality and providing practical policy advices. First, the higher the educational attainment, the less important the role of labor market frictions in shaping between-group inequality. The between-group inequality captured by wage premiums is mostly driven by workers' heterogeneity, but the difference in productivity among different education groups is largely a consequence of a sorting mechanism of schooling rather than its productivity enhancement process. Second, for within-group inequality, the contribution of labor market frictions decreases with education. That is, the higher the educational attainment, the higher the dispersion of unobservable productivity differences among workers due to the sorting process. This suggests that a policy promoting educational attainment for reducing economic inequality might achieve the opposite result.

Our model environment can be enriched along several dimensions. First, the only way for the workers in our model to enhance their productivity is to pursue higher education. Although educational attainment is arguably the most important component among the observable characteristics that explain wage differentials, workers can in fact invest in their productivity while working, i.e., on-the-job training and learning-by-doing. This way of productivity enhancement, albeit it certainly makes the model much more complicated, may have an interesting interaction with labor market friction (see e.g., Bagger et al. (2014)). Second, our model features homogenous return-to-education for sake of simplicity. Assuming heterogeneous returns would contribute to larger within-group wage dispersion. We leave these extensions for future work.

References

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002, March). Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999, March). High Wage Workers and High Wage Firms. *Econometrica 67*(2), 251–334.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science 344*(6186), 843–851.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008, May). Trends in U.S. Wage Inequality: Revising the Revisionists. *The Review of Economics and Statistics 90*(2), 300–323.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014, June). Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics. *American Economic Review 104*(6), 1551–96.

Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review 39*(2), 257–73.

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

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality*. *The Quarterly Journal of Economics 128*(3), 967–1015.

Eckstein, Z. and G. J. van den Berg (2007, February). Empirical labor search: A survey. *Journal of Econometrics 136*(2), 531–564.

Flinn, C. and J. Mullins (2015). Labor market search and schooling investment. *International Economic Review 50*, 359–398.

Goldin, C. and L. F. Katz (2007). Long-Run Changes in the Wage Structure: Narrowing, Widening, Polarizing. *Brookings Papers on Economic Activity 38*(2), 135–168.

Huggett, M., G. Ventura, and A. Yaron (2011). Sources of lifetime inequality. *The American Economic Review 101*(7), pp. 2923–2954.

Kremer, M. and E. Maskin (1996, August). Wage inequality and segregation by skill. Working Paper 5718, National Bureau of Economic Research.

Krueger, D. and F. Perri (2006). Does Income Inequality Lead to Consumption Inequality? Evidence and Theory. *Review of Economic Studies 73*(1), 163–193.

Krusell, P., L. E. Ohanian, J.-V. RÌos-Rull, and G. L. Violante (2000, September). Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica 68*(5), 1029–1054.

Lee, D., S. Y. Lee, and Y. Shin (2015). The option value of human capital: Higher education and wage inequality. mimeo.

Lemieux, T. (2006). Postsecondary Education and Increasing Wage Inequality. *American Economic Review 96*(2), 195–199.

Lopes de Melo, R. (2015). Firm wage differentials and labor market sorting: Reconciling theory and evidence. Technical report.

Mortensen, D. T. (2003). *Wage Dispersion: Why Are Similar Workers Paid Differently?*, Volume 1 of *MIT Press Books*. The MIT Press.

Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium Wage Dispersion with Worker and Employer Heterogeneity. *Econometrica 70*(6), 2295–2350.

Rosenzweig, M. R. and K. I. Wolpin (2000, December). Natural "Natural Experiments" in Economics. *Journal of Economic Literature 38*(4), 827–874.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2015, May). Firming up inequality. Working Paper 21199, National Bureau of Economic Research.

Stange, K. M. (2012). An Empirical Investigation of the Option Value of College Enrollment. *American Economic Journal: Applied Economics 4*(1), 49–84.

Trachter, N. (2015, 03). Stepping stone and option value in a model of postsecondary education. *Quantitative Economics 6*(1), 223–256.

Violante, G. L. (2008). skill-biased technical change. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Basingstoke: Palgrave Macmillan.

A Omitted Proofs

Proof of Proposition 1. Proving linearity is straightforward, since it only requires to guess and verify. Thus, profit per worker in Equation 6 can be expressed as:

$$(\rho + \delta_i + \lambda_i^E(1 - F^i(q(w,p))))\pi^i(w,p) = \rho(p - w) + \lambda_i^E \int_{q(w,p)}^{p} \pi^i(\phi(x,p),p)dF^i(x)$$

Integrating by parts the last expression in the right hand side yields:

$$(\rho + \delta_i + \lambda_i^E)\pi^i(w,p) = \rho(p - w) - \lambda_i^E \int_{q(w,p)}^{p} \pi_w^i(\phi(x,p),p)\phi_1(x,p)F^i(x)dx \tag{44}$$

Differentiating both sides of this expression with respect to w yields:

$$(\rho + \delta_i + \lambda_i^E)\pi_w^i(w,p) = -\rho + \lambda_i^E \phi_1(q(w,p),p)q_w(w,p)\pi_w^i(w,p)F^i(q(w,p))$$

Now, notice that the definition of $q(w,p)$ implies that $\phi_1(p',p)q_w(w,p) = 1$, which then allows us to write:

$$\pi_w^i(w,p) = -\frac{\rho}{\rho + \delta_i + \lambda_i^E(1 - F^i(q(w,p)))} \tag{45}$$

We plug in this expression back in 44 to obtain:

$$(\rho + \delta_i + \lambda_i^E)\pi^i(w,p) = \rho(p - w) + \rho\lambda_i^E \int_{q(w,p)}^{p} \frac{\phi_1(x,p)}{\rho + \delta_i + \lambda_i^E(1 - F^i(x))} F^i(x)dx$$

We can manipulate the first term in the right hand side and write:

$$(\rho + \delta_i + \lambda_i^E)\pi^i(w,p) = \rho \int_{q(w,p)}^{p} \phi_1(x,p)dx + \rho\lambda_i^E \int_{q(w,p)}^{p} \frac{\phi_1(x,p)}{\rho + \delta_i + \lambda_i^E(1 - F^i(x))} F^i(x)dx$$

which simplifies to:

$$\pi^i(w,p) = \int_{q(w,p)}^{p} \frac{\rho}{\rho + \delta_i + \lambda_i^E(1 - F^i(x))} \phi_1(x,p)dx \tag{46}$$

From (18), the derivative of the wage equation with respect to the first argument is:

$$\phi_1(p',p) = (1 - \beta_i)\frac{\rho + \delta_i + \lambda_i^E(1 - F^i(p'))}{\rho + \delta_i + \lambda_i^E \beta_i(1 - F^i(p'))} \tag{47}$$

Plugging this into (46) yields (20). ■

Proof of Proposition 2. Linearity of U^i requires linearity of W^i. In both cases, this property can be easily verified using (9) and (10). Following Cahuc et al. (2006), when an unemployed is matched with type-p firm, he is offered $\phi_0(p)$ and obtains

$$W^i(\phi_0(p), p) = U^i + \beta_i \left(W^i(p,p) - U^i\right) \tag{48}$$

When a worker employed in a type-p firm receives an offer from a type-p' poacher with $p' > p$, he switches jobs and obtains

$$W^i(\phi(p,p'), p') = W^i(p,p) + \beta_i \left(W^i(p',p') - W^i(p,p)\right) \tag{49}$$

or alternatively

$$W^i(w,p) = W^i(q(w,p), q(w,p)) + \beta_i \left(W^i(p,p) - W^i(q(w,p), q(w,p))\right) \tag{50}$$

where we have just plugged in the definition of $q(w,p)$ in (49). The recursive expression for the value of being unemployed is given by:

$$\rho U_i = \rho b_i + \lambda_i^U \int_b^{\overline{p}} \left(W(\phi_0(x), x) - U_i\right) dF^i(x) \tag{51}$$

Using the expression for $W^i(\phi_0(x), x)$ in (48) and rearranging terms we obtain

$$(\rho + \lambda_i^U \beta_i) U^i = \rho b_i + \lambda_i^U \beta_i \int_b^{\overline{p}} W^i(x,x) dF^i(x) \tag{52}$$

To obtain a sharper characterization of this function we work first with the value of being employed. The value of being employed at wage w within a type-p firm is given by:

$$\begin{aligned}(\rho + \delta_i + \lambda_i^E(1 - F^i(q(w,p))))W^i(w,p) = \quad & \rho w + \delta_i U^i + \lambda_i^E \int_{q(w,p)}^{p} W^i(\phi(x,p), p) dF^i(x) \\ & + \lambda_i^E \int_p^{\overline{p}} W^i(\phi(p,x), x) dF^i(x)\end{aligned} \tag{53}$$

We now consider the case $w = p$. Notice this implies $q(p,p) = p$. Using (49) we obtain:

$$(\rho + \delta_i + \lambda_i^E \beta_i (1 - F^i(p)))W^i(p,p) = \rho p + \delta_i U^i + \lambda_i^E \beta_i \int_p^{\overline{p}} W^i(x,x) dF^i(x) \tag{54}$$

and then using (52) we obtain:

$$\begin{aligned}(\rho+\delta_i+\lambda_i^E\beta_i(1-F^i(p)))W^i(p,p) = \quad & \rho(p-b)+(\rho+\delta_i+\lambda_i^U\beta_i)U^i \\ & -\lambda_i^U\beta_i\int_b^{\overline{p}}W^i(x,x)dF^i(x)+\lambda_i^E\beta_i\int_p^{\overline{p}}W^i(x,x)dF^i(x)\end{aligned}$$

Integrating by parts we obtain:

$$\begin{aligned}(\rho+\delta_i+\lambda_i^E\beta_i)W^i(p,p) = \quad & \rho(p-b)+(\rho+\delta_i+\lambda_i^U\beta_i)U^i-\lambda_i^U\beta_i\left(W^i(\overline{p},\overline{p})-\int_b^{\overline{p}}W_p^i(x,x)F^i(x)dx\right) \\ & +\lambda_i^E\beta_i\left(W^i(\overline{p},\overline{p})-\int_p^{\overline{p}}W_p^i(x,x)F^i(x)dx\right)\end{aligned}$$

Taking derivatives to both sides of the last equation delivers an expression for the derivative of W^i with respect to the second argument p which is given by:

$$W_p^i(p,p)=\frac{\rho}{\rho+\delta_i+\lambda_i^E\beta_i(1-F(p))}$$

Plugging this back and rearranging terms yields:

$$\begin{aligned}(\rho+\delta_i+\lambda_i^E\beta_i)W^i(p,p) = \quad & \rho(p-b)+(\rho+\delta_i+\lambda_i^U\beta_i)U^i+(\lambda_i^E-\lambda_i^U)\beta_iW^i(\overline{p},\overline{p}) \\ & +\lambda_i^U\beta_i\int_b^{\overline{p}}\frac{\rho}{\rho+\delta_i+\lambda_i^E\beta_i(1-F(x))}F^i(x)dx \\ & -\lambda_i^E\beta_i\int_p^{\overline{p}}\frac{\rho}{\rho+\delta_i+\lambda_i^E\beta_i(1-F(x))}F^i(x)dx\end{aligned}$$

Notice that making $p=\overline{p}$, we obtain an expression for $W^i(\overline{p},\overline{p})$ given by:

$$W^i(\overline{p},\overline{p})=U^i+\frac{\rho}{\rho+\delta_i+\lambda_i^U\beta_i}\int_b^{\overline{p}}\frac{\rho+\delta_i+\lambda_i^E\beta_i(1-F^i(x))+\lambda_i^U\beta_iF^i(x)}{\rho+\delta_i+\lambda_i^E\beta_i(1-F^i(x))}dx$$

Again, plugging this back and rearranging terms we obtain:

$$\begin{aligned}W^i(p,p)=U^i \quad & + \quad \frac{\rho}{\rho+\delta_i+\lambda_i^U}\left(\int_b^p\frac{\rho+\delta_i+\lambda_i^E\beta_i(1-F^i(x))+\lambda_i^U\beta_iF^i(x)}{\rho+\delta_i+\lambda_i^E\beta_i(1-F^i(x))}dx\right. \\ & - \quad \left.\int_p^{\overline{p}}\frac{(\lambda_i^U-\lambda_i^E)(1-F^i(x))}{\rho+\delta_i+\lambda_i^E\beta_i(1-F^i(x))}dx\right)\end{aligned}$$

Using this equation in (52) we obtain:

$$
\begin{aligned}
U_i &= b + \frac{\lambda_i^U \beta_i}{\rho + \delta_i + \lambda_i^U} \int_b^{\overline{p}} \left(\int_b^y \frac{\rho + \delta_i + \lambda_i^E \beta_i (1 - F^i(x)) + \lambda_i^U \beta_i F^i(x)}{\rho + \delta_i + \lambda_i^E \beta_i (1 - F^i(x))} dx \right. \\
&- \left. \int_y^{\overline{p}} \frac{(\lambda_i^U - \lambda_i^E)(1 - F^i(x))}{\rho + \delta_i + \lambda_i^E \beta_i (1 - F^i(x))} dx \right) dF(y)
\end{aligned}
$$

Finally, integrating by parts the second term we obtain

$$
U_i = b + \int_b^{\overline{p}} \frac{\lambda_i^U \beta_i (1 - F^i(x))}{\rho + \delta_i + \lambda_i^E \beta_i (1 - F^i(x))} dx
$$

which corresponds to (26). ■

Proof of Proposition 3. Plugging (27) and (28) into (29) and (30), we obtain

$$
\begin{aligned}
\rho U^{i-1} - (1 - c_i) b &= \theta^i(\overline{q}^i) \left[(1 + \varphi_i) S^{i+1}(\hat{q}(\overline{q}^i)) - U^{i-1} \right] \\
\rho U^{I-1} - (1 - c_I) b &= \theta^I(\overline{q}^I) \left[(1 + \varphi_I) U^I - U^{I-1}) \right]
\end{aligned}
$$

The right hand sides of these equations represent the expected gains of staying at school for one more instant, which must equal the opportunity cost of doing so, the left hand sides. Rearranging terms we get (31) and (32). ■

B Average unemployment duration

Conditional on being unemployed, if an agent retires/dies in period τ, unemployment duration is given by:

$$
\int_0^\tau t \lambda_i e^{-\lambda_i t} dt
$$

which can be written as

$$
-\tau e^{-\lambda_i \tau} + \int_0^\tau e^{-\lambda_i t} dt
$$

Solving the integral in the second term yields:

$$
-\tau e^{-\lambda_i \tau} - \frac{e^{-\lambda_i \tau} - 1}{\lambda_i}
$$

This is the expected unemployment duration for an agent that retires or dies at τ. Notice that as $\tau \to \infty$, this expression tends to $1/\lambda_i$ as usual. Aggregating over all agents we obtain:

$$\begin{aligned}\int_0^\infty \left(\frac{1-e^{-\lambda_i\tau}}{\lambda_i} - \tau e^{-\lambda_i\tau}\right)\mu e^{-\mu\tau}d\tau &= \frac{1}{\lambda_i} - \frac{\mu}{\lambda_i(\mu+\lambda_i)} - \frac{\mu}{(\mu+\lambda_i)^2} \\ &= \frac{(\lambda_i+\mu)^2 - \mu(\lambda_i+\mu) - \mu\lambda_i}{\lambda_i(\mu+\lambda_i)^2} \\ &= \frac{\lambda_i}{(\mu+\lambda_i)^2}\end{aligned}$$

Observe that when $\mu = 0$, then unemployment duration equals $1/\lambda_i$, which is the standard result.

C Stationary Distribution of Educational Attainment

In the model with $I = 2$, workers can be in one of the following five states: no breakthroughs, one breakthrough, high school graduate, college dropout, and college graduate (two breakthroughs). In this appendix we provide closed form expressions for the mass of workers in each state.

No breakthroughs

The mass of workers of ability $j \in \{0, 1\}$ with no breakthroughs can be expressed as follows

$$\Psi_j^1 = \mu \int_0^{t(\overline{q}^1, q_0)} e^{-(\mu+\theta_j^1)s} ds$$

where $t(\overline{q}^1, q_0)$ denotes the time at which worker decides to dropout. Since beliefs depreciate continuously according to (1), we have that in the first stage

$$q(t) = \frac{q_0}{(1-q_0)e^{\Delta\theta^1 t} + q_0}$$

Therefore,

$$t(\overline{q}^1, q_0) = \frac{1}{\Delta\theta^1} \ln\left(\frac{q_0(1-\overline{q}_1)}{\overline{q}_1(1-q_0)}\right)$$

So we obtain that

$$\Psi_j^1 = \frac{\mu}{\mu+\theta_j^1}(1 - e^{-(\mu+\theta_j^1)t(\overline{q}^1, q_0)}) \quad (55)$$

One breakthrough

Let $t^\star(\overline{q}^1, \overline{q}^2, q_0) = \min\{t(\overline{q}^1, q_0), \hat{q}^{-1}(\overline{q}^2)\}$. The set $\{\hat{q}(q(t)) \mid t \in [q_0, t^\star(\overline{q}^1, \overline{q}^2, q_0))\}$ contains all possible beliefs of agents with one breakthrough. The mass of workers of ability $j \in \{0, 1\}$ with one breakthrough can be expressed as follows:

$$\Psi_j^2 = \mu\theta_j^1 \int_0^{t^\star(\overline{q}^1,\overline{q}^2,q_0)} e^{-(\mu+\theta_j^1)s} \left(\int_0^{t(\overline{q}_2,\hat{q}(q(s)))} e^{-(\mu+\theta_j^2)\tau} d\tau \right) ds$$

We need to distinguish two situations. If $\overline{q}^2 = 0$, then:

$$\Psi_j^2 = \frac{\mu\theta_j^1}{(\mu+\theta_j^1)(\mu+\theta_j^2)}(1 - e^{-(\mu+\theta_j^1)t^\star(\overline{q}^1,\overline{q}^2,q_0)}) \tag{56}$$

Otherwise we have:

$$\Psi_j^2 = \frac{\mu\theta_j^1}{\mu+\theta_j^2} \int_0^{t^\star(\overline{q}^1,\overline{q}^2,q_0)} e^{-(\mu+\theta_j^1)s} \left(1 - e^{-(\mu+\theta_j^2)t(\overline{q}_2,\hat{q}(q(s)))} \right) ds$$

It is possible to write:

$$e^{-(\mu+\theta_j^2)t(\overline{q}_2,\hat{q}(q(s)))} = (e^{\Delta\theta^2 t(\overline{q}_2,\hat{q}(q(s)))})^{-\frac{\mu+\theta_j^2}{\Delta\theta^2}} = \left(\frac{\theta_0^1}{\theta_1^1} \frac{(1-q_0)\overline{q}_2}{(1-\overline{q}_2)q_0} \right)^{\frac{\mu+\theta_j^2}{\Delta\theta^2}} e^{\frac{\Delta\theta^1}{\Delta\theta^2}(\mu+\theta_j^2)s}$$

Hence,

$$\Psi_j^2 = \frac{\mu\theta_j^1}{\mu+\theta_j^2} \int_0^{t^\star(\overline{q}^1,\overline{q}^2,q_0)} \left(e^{-(\mu+\theta_j^1)s} - \left(\frac{\theta_0^1}{\theta_1^1} \frac{(1-q_0)\overline{q}_2}{(1-\overline{q}_2)q_0} \right)^{\frac{\mu+\theta_j^2}{\Delta\theta^2}} e^{-(\mu+\theta_j^1 - \frac{\Delta\theta^1}{\Delta\theta^2}(\mu+\theta_j^2))s} \right) ds$$

We must entertain the possibility that $\mu + \theta_j^1 - \frac{\Delta\theta^1}{\Delta\theta^2}(\mu + \theta_j^2) = 0$, in which case we obtain

$$\Psi_j^2 = \frac{\mu\theta_j^1}{(\mu+\theta_j^2)(\mu+\theta_j^1)}(1 - e^{-(\mu+\theta_j^1)t^\star(\overline{q}^1,\overline{q}^2,q_0)}) - \frac{\mu\theta_j^1 \left(\frac{\theta_0^1}{\theta_1^1} \frac{(1-q_0)\overline{q}_2}{(1-\overline{q}_2)q_0} \right)^{\frac{\mu+\theta_j^2}{\Delta\theta^2}}}{\mu+\theta_j^2} t^\star(\overline{q}^1, \overline{q}^2, q_0) \tag{57}$$

Otherwise:

$$\Psi_j^2 = \frac{\mu\theta_j^1}{(\mu+\theta_j^2)(\mu+\theta_j^1)}(1 - e^{-(\mu+\theta_j^1)t^\star(\overline{q}^1,\overline{q}^2,q_0)}) - \frac{\mu\theta_j^1 \left(\frac{\theta_0^1}{\theta_1^1}\frac{(1-q_0)\overline{q}_2}{(1-\overline{q}_2)q_0}\right)^{\frac{\mu+\theta_j^2}{\Delta\theta^2}}}{(\mu+\theta_j^2)(\mu+\theta_j^1 - \frac{\Delta\theta_j^1}{\Delta\theta_j^2}(\mu+\theta_j^2))}\left(1 - e^{-\left(\mu+\theta_j^1-\frac{\Delta\theta_j^1}{\Delta\theta_j^2}(\mu+\theta_j^2)\right)t^\star(\overline{q}^1,\overline{q}^2,q_0)}\right) \tag{58}$$

High School Graduate

The mass of workers of ability $j \in \{0,1\}$ that become high school graduates can be expressed as follows:

$$\Psi_j^3 = \mu e^{-(\mu+\theta_j^1)t(\overline{q}_1,q_0)}\int_0^\infty e^{-\mu\tau}d\tau = e^{-(\mu+\theta_j^1)t(\overline{q}_1,q_0)} \tag{59}$$

College Dropout

The mass of workers of ability $j \in \{0,1\}$ that become dropouts can be expressed as follows:

$$\Psi_j^4 = \mu\theta_j^1\int_0^{t^\star(\overline{q}^1,\overline{q}^2,q_0)} e^{-(\mu+\theta_j^1)s}\left(e^{-(\mu+\theta_j^2)t(\overline{q}_2,\hat{q}(q(s)))}\int_0^\infty e^{\mu\tau}d\tau\right)ds + \mu\theta_j^1\int_{t^\star(\overline{q}^1,\overline{q}^2,q_0)}^{t(\overline{q}^1,q_0)} e^{-(\mu+\theta_j^1)s}\left(\int_0^\infty e^{\mu\tau}d\tau\right)ds$$

Again, we must entertain two cases. If $\mu+\theta_j^1 - \frac{\Delta\theta^1}{\Delta\theta^2}(\mu+\theta_j^2) = 0$, we obtain:

$$\Psi_j^4 = \theta_j^1\left(\frac{\theta_0^1}{\theta_1^1}\frac{(1-q_0)\overline{q}_2}{(1-\overline{q}_2)q_0}\right)^{\frac{\mu+\theta_j^2}{\Delta\theta^2}} t^\star(\overline{q}^1,\overline{q}^2,q_0) + \frac{\theta_j^1}{\mu+\theta_j^1}\left(e^{-(\mu+\theta_j^1)t^\star(\overline{q}^1,\overline{q}^2,q_0)} - e^{-(\mu+\theta_j^1)t(\overline{q}^1,q_0)}\right) \tag{60}$$

Otherwise

$$\Psi_j^4 = \frac{\theta_j^1\left(\frac{\theta_0^1}{\theta_1^1}\frac{(1-q_0)\overline{q}_2}{(1-\overline{q}_2)q_0}\right)^{\frac{\mu+\theta_j^2}{\Delta\theta^2}}}{\mu+\theta_j^1-\frac{\Delta\theta_j^1}{\Delta\theta_j^2}(\mu+\theta_j^2)}\left(1 - e^{-\left(\mu+\theta_j^1-\frac{\Delta\theta_j^1}{\Delta\theta_j^2}(\mu+\theta_j^2)\right)t^\star(\overline{q}^1,\overline{q}^2,q_0)}\right) + \frac{\theta_j^1}{\mu+\theta_j^1}\left(e^{-(\mu+\theta_j^1)t^\star(\overline{q}^1,\overline{q}^2,q_0)} - e^{-(\mu+\theta_j^1)t(\overline{q}^1,q_0)}\right) \tag{61}$$

College Graduate (two breakthroughs)

The mass of workers of ability $j \in \{0,1\}$ that become college graduates can be expressed as follows:

$$\Psi_j^5 = \theta_j^2 \Psi_j^2 \int_0^\infty e^{-\mu\tau} d\tau = \frac{\theta_j^2}{\mu} \Psi_j^2 \tag{62}$$

Given a pair of values for the belief cutoffs $(\overline{q}^1, \overline{q}^2)$, equations (55) - (62) allows us to calculate the stationary distribution in closed form. These expressions will prove very useful when targeting the workers' distribution observed in the data.

D Calibration Details

We use data on the schooling attainment distribution and labor market data on unemployment rate and unemployment duration.

- Given a value of μ and unemployment duration data τ_i^{data}, Equation 43 pins down the equilibrium contact rate for each submarket, which in this section we denote λ_i^{data}.
- Given (μ, λ_i^{data}) we use Equation 17 and unemployment rate data u_i^{data} to recover the separation rates for each submarket, which we denote δ_i^{data}
- Let Ψ_i^{data} denote the mass of workers participating in submarket i in the data. Proposition 3 shows that the cutoffs $\overline{q}^i$ do not depend on the initial productivity z_0. Therefore, given an initial distribution of beliefs q_0, it is possible to calculate the belief cutoffs that deliver Ψ_i^{data} by following Appendix C.

Using 12 and aggregating 23 across all firms we obtain:

$$\frac{\lambda_i^{data} \Psi_i^{data}}{m} = \frac{\int_b^{\overline{p}} \overline{\pi}_i(p) d\Gamma(p) \mathrm{E}[z \mid i]}{\chi_i} \tag{63}$$

Hence, in order to match the data, we can use χ_i/m or β_i which relates negatively with firms' profits. We have less control over the conditional expectations because, as we will see, they need to be consistent with the equilibrium unemployment values. Notice that for any two educational levels

$$\frac{\lambda_i^{data}}{\lambda_i^{data}} \frac{\Psi_i^{data}}{\Psi_j^{data}} = \frac{\int_b^{\overline{p}} \overline{\pi}_i(p) d\Gamma(p)}{\int_b^{\overline{p}} \overline{\pi}_j(p) d\Gamma(p)} \frac{\mathrm{E}[z \mid i]}{\mathrm{E}[z \mid j]} \tag{64}$$

In the data, unemployment duration does not depend on educational attainment. Also, there are more HS than CG and more CG than DO. In other words, if we compare $i = HS$ and $j = DO$, the LHS of this expression is greater than 1. Assuming that conditional expectations are increasing in educational attainment, a necessary condition for the model to be consistent with the data is then $\beta_{HS} < \beta_{DO}$, since this will guarantee that profits for the firm in the HS market are higher.

On the other hand if we compare $i = DO$ and $j = CG$, then the LHS expression is less than 1. Again, assuming that conditional expectations are increasing, it turns out that $\beta_{DO} < \beta_{CG}$ is not necessary for the model to be consistent with the data unless the productivity differences between college graduates and dropouts are too large.

In other words, in a reasonable calibration we should obtain that the bargaining power parameter is increasing in schooling attainment. This also implies that the frictional component of wage dispersion within groups is decreasing in educational attainment. This frictional dispersion appears to be also convex in the bargaining power, which means that the level of β_1 matters.