

Frictional Labor Markets, Education Choices and Wage Inequality ^{*}

Manuel Macera [†]

Hitoshi Tsujiyama [‡]

November 2017

Abstract

This paper studies how education choices and labor market frictions interact in shaping wage inequality. The wage premium of college graduates relative to high school graduates (between-group inequality) has tripled since 1980 in the U.S., and the variance of log wages conditional on educational attainments (within-group inequality) has become about 50% larger across the board. To understand the source of this change, we construct a model with schooling investments and labor market frictions that generates supply and demand of skills and frictional wage differentials as equilibrium objects. The model features a two-sided sorting: education sorting of skilled workers into college education and labor market sorting of productive firms into the labor market for college graduates – together implying an assortative matching of high skilled workers to productive firms. A novel model-based wage decomposition of both the between- and within-group inequalities is obtained. Calibrating the model to the U.S. data, we find that the inequality trend is accounted for by worker composition and labor market friction. If there were no skill-biased technological change, the variance of log wages would be smaller, mainly due to lower within-group inequality.

^{*}We thank Naoki Aizawa, Cristina Arellano, Nicola Fuchs-Schündeln, Jonathan Heathcote, Loukas Karabarbounis, Leo Kaas, John Kennan, Zachary Mahone, Andreas Mueller, Roberto Pinheiro, Fabrizio Perri, Ctirad Slavik, Naoki Takayama, Satoshi Tanaka and Gianluca Violante for useful comments and discussions. All errors remain our own.

[†]Universidad Torcuato Di Tella. mmacera@utdt.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 data sets. 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 wages¹. 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 studies 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

¹Unobserved 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 2005². 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

²To 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.

- 2006 Education and Wage Inequality, Book finds in most European countries the observed rise in earnings inequality was driven primarily by rises in inequality within (rather than between) groups of workers and education is the most relevant factor (rather than age, sex or sector of employment).

Related Literature

[To be completed] 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 search³. 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 interac-

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

tion 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 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. \(2017\)](#)). 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.

contribution: the determinants of residual inequality

- Chade and Lindenlaub (2017) analyze how a change in risks affects agent's risky pre-match investment decisions when the matching is frictionless.
- A number of papers emphasize the role of supply and demand of skills in explaining the inequality trends (e.g., [Goldin and Katz \(2007\)](#), [Autor \(2014\)](#)).
- Autor RES
- Acemoglu (1998, QJE), a model with endogenous technology upgrading. An increase in the supply of skills increases the market size for skill-complementary technologies and may induce skill-bias technical change, which can account for both short- and long-run trend of the college premium during 1970s and 80s.
- Acemoglu (1999, AER), a model with endogenous job creation. An increase in the supply of skills changes firms from creating a single type of job for all workers to creating different type of jobs for different skills.
- Acemoglu (2002, JEL) An acceleration in skill bias during the past few decades appears to be the main cause of the increase in inequality.
- Taber (2001) argues that increase in demand for skills leading a rise in the variance of

wage residuals could at the same time drive an increase in the college premium. Card and Lemieux (1993). Lemieux, T. (2010). LLS

- While a number of studies find that the expansion of education led to a modest reduction in the college wage premium (Juhn et al., 2005; Carneiro and Lee, 2011), other studies infer a sizable increase (Kaymak, 2009; Bowlus and Robinson, 2012).
- Hendricks Schoellman JME
- dynamic human capital models with education choice (Keane and Wolpin, 1997; Lee, 2005; Lee and Wolpin, 2006): no uncertainty in education and frictionless labor markets
- Search/matching and education: Holzner Launov (2010, EER) estimate social and private returns to education in Germany using a labor-search model with a perfect education sorting and a wage posting game. Acemoglu (1996): education, no labor search, random matching. Many firms hoping to match with the more skilled workers will invest more. Therefore, even some of the workers who have not increased their human capital-and who are competing for the same jobs-end up working with more physical capital and earning an increased rate of return on their human capital. In other words, as recently emphasized by Lucas [1988], the rate of return on human capital of a worker is increasing in the human capital stock of the workforce. - Rent sharing translates into a hold-up problem for the workers. The existence of unemployment leads to underinvestment in education. Their reasoning is based on a hold-up argument: the existence of unemployment and search costs creates room for wage bargaining and rent sharing between worker and employer. The more productive the worker, the more rent there will be to share. This then means that a worker who invests in education receives less than a full share of the return. There is a positive externality from education on firms, and as a result there is underinvestment in education. Moen (1999): analyze the incentives for workers to invest in education when there is an unemployment risk. Show a potential of overturning the result in Acemoglu 96 with a rat race story. Assume that one part of the private gains from education may be that it reduces the probability of being unemployed (the arrival rate of job offers is independent of a worker's education) and show that overinvestment in human capital compared to the socially optimal level is possible. investing in education improves one's ranking in the job queue, but at the expense of the others. FlinnMullins.

[assortative matching, theory] In the context of wage inequality, assortative matching of high skilled workers to productive firms is particularly relevant. Guvenen et al (2016) find that the assortative matching can account for the half of the change in between-firm wage

variation that is a major cause of increasing wage inequality over the last 40 years.⁴ There has been important theoretical attempts for generating an assortative matching in search models. Shimer and Smith (2000) characterize the conditions on match output to obtain sorting in the partnership models in which search is precluded once in a match. In the on-the-job search models, Bagger et al (2014) assume ability-dependent job finding rates for workers, and Bagger and Lentz (2017) endogenize this feature as an equilibrium outcome by introducing endogenous search intensity. With this way, more productive workers climb the job ladder faster and hence tend to work at more productive firms.⁵ In our model, sorting obtains by a different mechanism through education choices and firms' sorting into labor markets. This mechanism is also empirically supported: for example, using matched employer-employee data, Engbom and Moser (2017) show that higher education degrees help sorting towards high wage firms and this sorting explains substantial part of the return to college.

[XX] Robin RED. positive selection into college enrollment. Heckman et al. (2006)

2 Trends in Wage Inequality

This section provides the empirical evidence of the trends in inequality in the U.S. Our main focus is on the standard measures of wage inequality used in the literature (e.g. Card et al. 2013), namely, between- and within-group wage inequality. Our sample is full-time working white males aged from 25-55 in the Current Population Survey (CPS). A more detailed data description is found in Appendix A. The moments are five-year time-averaged to mitigate the cyclical effects (e.g, the 1990 value is the average of 1988-1992).

Between-Group Inequality Trend Panel A of Figure 1 plots the trend of between-education-group inequality measured by the ratio of the average wage of college graduates / some college attendees to that of high school graduates. This measure, also known as wage premium, 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 much milder and has stopped in the 1990s.

⁴Positive assortative matching is found for various countries, e.g., Brasil (Lopes de Melo, 2013), Denmark (Bagger and Lentz, 2017), Germany (Card et al, 2013; Hagedorn et al., 2017) and Sweden (Bonhomme et al., 2015).

⁵Another example of sorting in the job search model is Helpman Itzhaki Redding (2010) that construct a model in which firms can screen workers to improve the workers composition. Because more productive firms have higher returns to screening, they screen more intensively and thus have workforces of higher average ability than less productive firms.

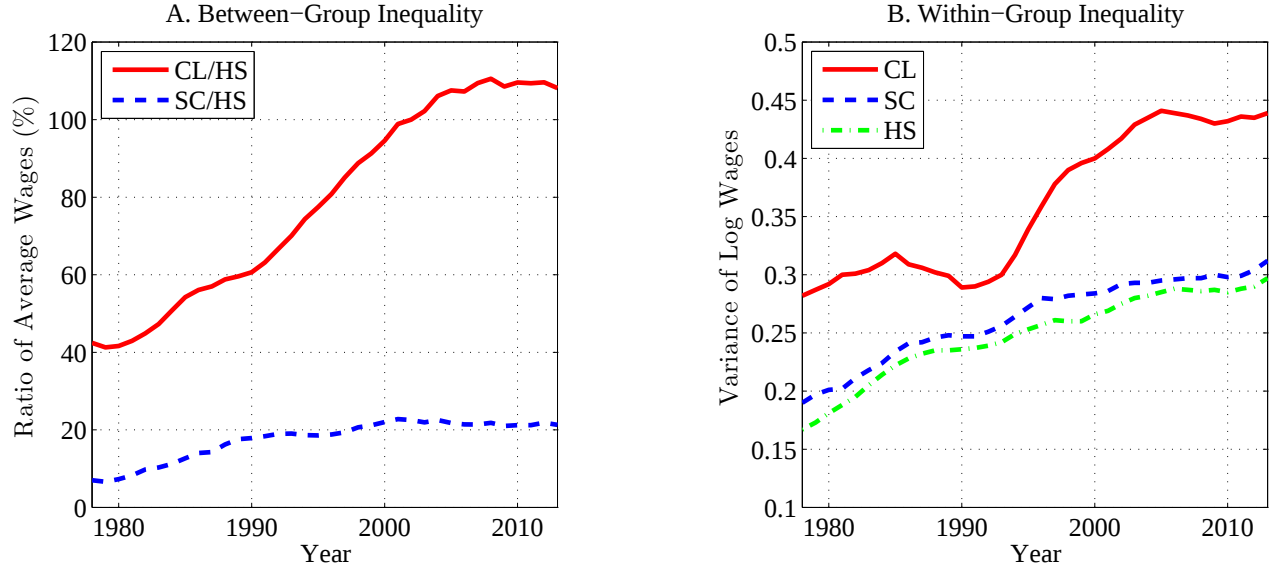


Figure 1: Between- and Within-Group Wage Inequality. Panel A plots the between-group inequality measured by the ratio of average wage of college graduates (CL) / some college (SC) attendees to that of high school graduates (HS). Panel B plots the within-group inequality of CL, SC and HS measured by the variance of log wages. The values are five-year time-averaged.

Within-Group Inequality Trend Panel B of Figure 1 plots the trend of within-education-group inequality measured by the variance of log wages for each education group, high school graduates (HS), some college attendees (SC), and college graduates (CL). The within-group inequality, also known as 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 around 0.45, and became flat again in the late 2000s. Overall, the variance of log wages has become one and half times larger over the past 30 years for all the three education groups.

Between-/Within-Firm Inequality Trend An important empirical fact regarding the decomposition of the inequality is also documented in Song et al. (2015). Namely, they find that the within-firm wage variation contributes more to the level of the overall unconditional wage variance over the past three decades, while the between-firm wage variation is more responsible for explaining the increase in wage inequality. Our model also aims for this trend.

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. Define the effective discount factor $\rho \equiv r + \mu$. Workers' lifetime is divided in two stages: a schooling stage and a labor market stage. In the schooling stage, workers enhance their labor productivity 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 schooling ability $a \in \mathcal{A}$ to complete schools and a skill $z \in \mathcal{Z} \subseteq \mathbb{R}$. For simplicity, assume $\mathcal{A} = \{a_0, a_1\}$ with $a_0 < a_1$. Assume a and z are potentially correlated and time-invariant.

We model education as a costly, risky investment taken before entering the labor markets. Schooling levels are indexed by $s \in \mathcal{S} \equiv \{1, \dots, S\}$, each associated with a parameter θ_s . Education is risky, as in [Manski \(1989\)](#), because the completion occurs by a Poisson process with arrival rate $\theta_i^s \equiv \theta_s \cdot a_i$. The parameter θ_s measures how long it takes to complete schooling level s on average, whereas ability a indicates how likely the worker would indeed complete it. Education is also costly in that, while in school, workers incur an instantaneous marginal utility cost c . This represents all kinds of costs associated with education including financial costs and psychic costs.⁶

The gain from this risky investment is twofold. First, upon completion, workers' labor market productivity changes from $h_{s-1}(z)$ to $h_s(z)$, where $h_s : \mathcal{Z} \rightarrow \mathbb{R}_+$ for $s \in \mathcal{S}_0 \equiv \{0\} \cup \mathcal{S}$ is a strictly increasing function that translates skill z into efficiency units of labor. Second, the labor markets are assumed to be segmented by schooling levels; only workers with a college degree can thus find a job in college market. In [Section 3.4](#), we discuss the model assumptions and implications in great detail.

Workers Problem Workers know their skill z , but not their ability a . Each worker holds a prior belief q that he/she is a high ability type, i.e. $q = \Pr\{a = a_1\}$, and its initial prior is given conditional on the observed z . The prior is then updated based on Bayes' rule.⁷ It follows

⁶[Heckman et al. \(2006\)](#) provide the evidence on the importance of psychic costs for the schooling decision.

⁷[Manski \(1989\)](#) calls this process as "schooling as experimentation", although he does not have an explicit learning in his model. Learning about academic ability is shown to play a prominent role in college enrollment and continuation decisions ([Stange 2012](#)). [Stinebrickner and Stinebrickner \(2012\)](#) provide direct evidence that students' aptitudes are uncertain and gradually learned through the performance in school. [Hendricks et al. \(2017\)](#) also document that academic ability is more important predictors of who attended college after World War II than family income or socioeconomic status.

immediately that, if a worker with prior q completes schooling level s , the prior jumps to $\hat{q}(q)$ such that

$$\hat{q}(q) = \frac{\theta_1^s}{\theta^s(q)} q \quad \text{where} \quad \theta^s(q) \equiv q\theta_1^s + (1-q)\theta_0^s,$$

and, otherwise, it depreciates continuously:⁸

$$\dot{q} = -\Delta\theta^s(1-q)q \quad \text{where} \quad \Delta\theta^s \equiv \theta_1^s - \theta_0^s. \quad (1)$$

The individual state of a worker in the schooling stage is given by a triple (q, z, s) .⁹ Workers hold the option to dropout schooling level s and enter immediately the labor market that is indexed by $s-1$. 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 quit schooling. Their optimal decision takes the form of a cutoff belief $\bar{q}^s : \mathcal{Z} \rightarrow [0, 1]$ that will trigger dropout while pursuing schooling level s if and only if $q < \bar{q}^s(z)$. Importantly, the optimal schooling decision generates a non-degenerate distribution of workers over pairs (z, s) in the labor markets.

Let V^s be the value of pursuing schooling level s when the option to dropout is not exercised. This function is of class $\mathcal{C}^1$ and must satisfy the following stochastic partial differential equation:

$$\rho V^s(q, z) = \rho(1-c)b \cdot h_{s-1}(z) + V_q^s(q, z)\dot{q} + \theta^s(q) [V^{s+1}(\hat{q}(q), z) - V^s(q, z)], \quad (2)$$

where V_q^s denotes the partial derivative of V^s with respect to q . The first term on the right hand side is the instantaneous utility, which we normalize by discount factor ρ for convenience. The flow value of leisure is denoted by b , which is linear in efficiency units of labor, $h_{s-1}(z)$, as standard. The second term captures the change in the value of schooling as a consequence of the deterioration in the worker's belief. The last term represents the change in the expected value of schooling upon completion. In this case, the worker moves to the next schooling level $s+1$ with a new prior, $\hat{q}(q)$.

In the case of $s = S$, completion means that the worker enters the labor market, thus $V^{S+1}(\hat{q}(q), z) = U^S(z)$ in equation (2), where $U^s(z)$ denotes the value of being unemployed in

⁸To see how the belief depreciates, consider prior q_t at time t . The probability of the worker finishing the schooling over the next time interval dt is given by $\theta_a^s dt$. Thus, when she does not succeed, q_t becomes

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

Subtracting q_t from both sides, dividing by dt and then taking $dt \rightarrow 0$ yields equation (1).

⁹The initial prior belief, denoted by $q_0(z)$, corresponds to the subjective probability that each worker attaches to being type a_1 , which is calculated by Bayes' rule (see [Appendix D](#)). The individual initial state is then given by $(q_0(z), z, 1)$.

submarket s for a worker with skill z . Upon completion, schooling ability and hence subjective belief q becomes irrelevant.¹⁰

3.2 Labor Market Stage

Labor markets are indexed by the level of schooling attainment. In each labor submarket s , workers and firms are matched randomly and production takes place. The individual state of a worker upon entry to the labor market consists of a pair (z, s) . Assume a worker's productivity is perfectly observed by a firm only after matching. Denote by $\Psi(z, s)$ the fraction of workers that participate in submarket s and have skill level less than or equal to z . Also denote by $\Psi^s(z)$ the distribution of skill levels conditional on being in submarket s .

Firms are heterogeneous in their productivity level, $p \in \mathcal{P} \equiv [b, \bar{p}]$. The production technology is linear in efficiency units of labor; the marginal product of a firm with productivity p (henceforth, p -firm) of hiring an additional worker with type (z, s) (henceforth, (z, s) -worker) is given by $p \cdot h_s(z)$. The total output of p -firm is then equal to p times the sum of its employees' efficient units of labor.

Wage Determination Our wage bargaining framework is based on [Rubinstein \(1982\)](#) and closely follows [Cahuc et al. \(2006\)](#), which we describe here succinctly. Workers participating in submarket s sample job offers at a rate λ_s^U if unemployed and λ_s^E if employed. The type of the firm from which the offer originates is drawn from the sampling distribution F^s . Upon matching, bargaining occurs under complete information. Denote by β 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 can be either employed or unemployed before matched with a p -firm. If the worker is unemployed, denote by $\phi_0^s(p)$ the wage resulting from the bargaining process between the worker and the firm. If the worker is employed by a p' -firm, the matched p -firm can poach the worker if and only if $p > p'$. In the case of poaching, the bargained wage is denoted by $\phi^s(p', p)$, where the first argument represents the productivity of the firm that made the best outside offer (i.e., the productivity of the last employer) and the second argument is the productivity of the new employer.¹¹ The incumbent p' -firm can successfully deter poaching, if $p' \geq p$. A renegotiation takes place, if and only if $p \in (g^s(w, p'), p']$, and results in a wage raise from the current wage w to $\phi^s(p, p')$, where $g^s(w, p')$ is defined as the productivity level satisfying $\phi^s(g^s(w, p'), p') = w$, that is, the maximum productivity that does not trigger a wage raise.

¹⁰Our model can be extended to incorporate learning-by-doing on the job whose degree also depends on a . In such a case, a would be a relevant state variable in the labor market stage. We leave this extension as our future work.

¹¹In the class of labor search models with on-the-job search and wage renegotiation, wages generally depend on the individual history of past offers. In the present model, p' is a sufficient statistics for the history.

Firms A firm that wants to hire a worker with schooling level s must create a vacancy in submarket s . 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 an exogenous rate δ_s .

Let $\pi^s(w, p, z)$ denote the profit per worker for a p -firm when hiring a (z, s) -worker at wage w . This function can be expressed recursively as

$$\begin{aligned} \rho\pi^s(w, p, z) &= \rho(p - w) \cdot h_s(z) - [\delta_s + \lambda_s^E(1 - F^s(p))] \pi^s(w, p, z) \\ &\quad + \lambda_s^E \int_{g^s(w, p)}^p [\pi^s(\phi^s(x, p), p, z) - \pi^s(w, p, z)] dF^s(x). \end{aligned} \quad (3)$$

The first term on the right hand side is the (normalized) flow profit. The second term is the capital loss that stems from separation or poaching that occurs when the worker receives an offer from a firm with productivity higher than p . The third term is the capital loss that occurs when the incumbent employer deters poaching at the cost of a wage raise.

Since vacancies are posted before matching, the incentives to create a new vacancy depend on expected profits rather than on realized profits. Upon matching, there are three possible outcomes for the vacancy: filled by an unemployed worker, filled by a worker previously employed by a firm with lower productivity, or unfilled. Hence, the expected profit from posting a vacancy in submarket s , $\Pi^s(p)$, must satisfy the following recursive expression:

$$\begin{aligned} \rho\Pi^s(p) &= \rho\pi_0 + \frac{u_s}{\lambda_s^U} \left[\int_{z \in \mathcal{Z}} \pi^s(\phi_0^s(p), p, z) d\Psi^s(z) - \Pi^s(p) \right] \\ &\quad + \frac{1 - u_s}{\lambda_s^E} \left\{ \int_{z \in \mathcal{Z}} \left[\int_b^p \pi^s(\phi^s(x, p), p, z) l^s(x) dx \right] d\Psi^s(z) - \Pi^s(p) \right\}, \end{aligned} \quad (4)$$

where the parameter π_0 is the flow payoff (cost) of having a vacancy posted, u_s is the unemployment rate in submarket s , and $l^s(p)$ is the mass of workers employed at a p -firm. The second/third term on the right hand side represents the expected profit if the firm is matched with an unemployed/employed worker.

A p -firm takes as given $\{l^s, F^s, \Psi^s, u_s, \lambda_s^E, \lambda_s^U, \phi_0^s, \phi^s\}_{s \in \mathcal{S}_0}$, and chooses an optimal vacancy creation policy $v_s(p)$ in order to solve the following problem: for each s

$$\max_{v_s} [\Pi^s(p)v_s - \kappa(v_s)], \quad (5)$$

where $\kappa(\cdot)$ is the vacancy creation cost function, which is assumed to be strictly increasing and strictly convex. We assume $\pi_0 = 0$ and free entry with no entry costs.¹²

¹²This assumption guarantees that all firms will participate. If $\pi_0 > 0$, then, upon entry, some firms might optimally choose not to post vacancies in a particular submarket, which is unnecessary complication (see e.g. Cahuc et al. 2006). On the other hand, if one assumes strictly positive entry costs $E > 0$, the measure of

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

$$\begin{aligned} \rho W^s(w, p, z) &= \rho w \cdot h_s(z) + \delta_s [U^s(z) - W^s(w, p, z)] \\ &\quad + \lambda_s^E \int_{g^s(w, p)}^p [W^s(\phi^s(x, p), p, z) - W^s(w, p, z)] dF^s(x) \\ &\quad + \lambda_s^E \int_p^{\bar{p}} [W^s(\phi^s(p, x), x, z) - W^s(w, p, z)] dF^s(x), \end{aligned} \quad (6)$$

$$\rho U^s(z) = \rho b \cdot h_s(z) + \lambda_s^U \int_b^{\bar{p}} [W(\phi_0^s(x), x, z) - U^s(z)] dF^s(x). \quad (7)$$

The first term on the right hand side of equation (6) 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 raise by the current employer. The last term is the expected net gain from accepting a higher wage offer made by a better employer. Likewise, the second term on the right hand side of equation (7) is the expected net gain from accepting a new offer.

Sampling Distribution and Contact Rates We assume the probability of being matched with a p -firm in submarket s is proportional to the number of vacancies each firm posts in each submarket. Thus, the sampling distribution function must satisfy:

$$F^s(p) = \frac{\int_{x \leq p} v_s(x) d\Gamma(x)}{\int_{x \in \mathcal{P}} v_s(x) d\Gamma(x)} \quad (8)$$

where Γ represents firms' productivity distribution.

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

$$\lambda_s^U = \frac{\int_{x \in \mathcal{P}} v_s(x) d\Gamma(x)}{\int_{z \in \mathcal{Z}} d\Psi(z, s)} m. \quad (9)$$

The numerator represents the number of vacancies posted by all firms participating in submarket s , whereas the denominator represents the number of workers looking for a job in the same submarket. We consider a different search technology for employed workers, which is represented

participating firms would change. Suppose firms know their productivity p before entry. A p -firm will then enter the labor markets if and only if $\sum_{s \in \mathcal{S}_0} [\Pi^s(p) v_s(p) - \kappa(v_s(p))] \geq E$. In equilibrium, the last firm entering the markets has productivity $\tilde{p}$ for which this equation holds with equality. The measure of participating firms would then be given by $\tilde{m} = \int_{p \geq \tilde{p}} d\Gamma(p) < m$.

by $\lambda_s^E = (1 - \zeta)\lambda_s^U$ with $\zeta \geq 0$.

3.3 Stationary Equilibrium

A *stationary equilibrium* consists of wage functions $\{\phi_0^s, \phi^s\}$, value functions for the workers $\{V^s, W^s, U^s\}$ and the firms $\{\pi^s, \Pi^s\}$, policy functions for the workers $\{\bar{q}^s\}$ and the firms $\{v_s\}$, contact rates and sampling distributions $\{\lambda_s^U, \lambda_s^E, F^s\}$, and workers' distribution Ψ such that:

- (i) the wage functions solve the wage bargaining problem;
- (ii) the policy functions solve the schooling decision problem and the vacancy creation problem;
- (iii) the value functions satisfy their respective recursive equations;
- (iv) the contact rates and sampling distributions are consistent with individual choices; and
- (v) workers' distribution is stationary.

3.4 Discussion

In this section, we discuss the model assumptions and implications in great detail.

Model of Education Education models based on [Becker \(1964\)](#) assume that individuals can *choose* the optimal years of schooling to maximize the expected payoff.¹³ Motivated by the sizable dropout population in the U.S., recent dynamic schooling choice models explicitly differentiate the college enrollment decision from the graduation decision, yet abstract from uncertainty whether a program of schooling will be completed.¹⁴ Following [Manski \(1989\)](#), we further distinguish the continuation decision from graduation; in our model, pursuing a degree is discretionary, but obtaining a degree is uncertain.¹⁵ This distinction is empirically relevant for two reasons.¹⁶ First, in the survey data many individuals who plan to complete college indeed drop out ([Altonji 1993](#)). Second, if those who want, but fail, to complete college are systematically different from those who actually complete college in terms of labor productivity, then abstracting from the completion uncertainty results in a biased estimate of inequality decomposition.

We assume that skill z is observed, but ability a is not. The former assumption is only for simplicity, and, if workers know at least the expected value of z , the resulting education attainment decision is exactly the same as in our model. The latter assumption is made to have endogenous college dropouts. There are several advantages with this setting. First, we

¹³See e.g., [Keane and Wolpin \(1997\)](#), [Card \(2001\)](#), [Lee \(2005\)](#).

¹⁴E.g., [Stange \(2012\)](#), [Stinebrickner and Stinebrickner \(2012\)](#), [Lee et al. \(2017\)](#).

¹⁵The uncertainty in completion is different from the exogenous failure risk considered in [Athreya and Eberly \(2016\)](#), since in our model dropping out is still a voluntary decision. [Hendricks and Leukhina \(2017\)](#) also construct a model in which the college continuation decision depends on students' ability that affects the probability of earning credits.

¹⁶This is also theoretically motivated by [Manski \(1989\)](#) that argues that, for schooling to be completed, the student must not only decide that it is worthwhile to persist to graduation, but also be able to pass the courses. He then states that "(t)hus, completion has both exogenous and endogenous determinants" (p.305).

can explicitly differentiate productivity enhancement from selection, thus return to education from worker composition effect in the measured college premium. Second, we can solve for the workers' stationary distribution in closed form, which allows us to deal with heterogeneity of both workers and firms.

We allow ability a and skill z to be correlated. This means that the higher the observed skill, the higher the initial prior. Because the education decision problem boils down to an optimal stopping problem, the schooling in our model entails education sorting, i.e., high-skilled workers also tend to be highly educated. The degree of sorting is governed by the correlation between a and z . If they are not correlated, there is no sorting. In this case, high school graduates and college graduates have the same skill distribution; their average ex-post labor productivity differs only due to return to education. In contrast, if a and z are perfectly correlated and thus ability is also observed, then the model would deliver a perfect sorting, corresponding to the case of [Flinn and Mullins \(2015\)](#). Since education is no longer a risky investment, only high-skilled workers finish college, which splits the continuous support of initial productivity into disconnected intervals after the schooling stage.¹⁷ Our estimation in [Section 5](#), however, rejects the perfect correlation between skill and ability.

We also assume that return to education is common, rather than heterogenous, across agents. This is because heterogeneity in return to education cannot be separately identifiable from initial skill heterogeneity. However, we do have heterogeneity in return among ex-ante identical workers, since the completion of schooling occurs stochastically.

Segmented Labor Markets The labor markets are assumed to be segmented by schooling levels.¹⁸ This assumption aligns with education signaling models à la [Spence \(1973\)](#) in which the education attainment partially reveals workers' productivity. It works as a signal in our model too, because firms post vacancies, knowing the productivity distribution of workers in each submarket, before seeing the exact individual type upon matching. Education in our model is, however, not wasteful unlike signaling models.

We rule out the option of college graduates taking jobs that do not require a college degree (see, e.g., [Lee et al. 2017](#)), although whether this option is empirically relevant is not yet concluded in the literature. In the context of labor search, the presence of this option must imply a much shorter unemployment duration for college graduates than for high school graduates, since the former has a higher job finding rate. However, the duration in the data is quite similar

¹⁷This means that the most productive worker with a high school degree would be strictly less productive than the least productive worker with a college degree. This is not only unrealistic but also unappealing, because it means that, absent of any change in the initial distribution, an increase in the within-group inequality for college automatically implies a reduction in that for high school, which is inconsistent with the inequality trends seen in [Section 2](#).

¹⁸Like us, [van den Berg and Ridder \(1998\)](#) consider a wage posting model in which labor markets are segmented by observables such as the educational level.

for these two groups.¹⁹

Production Function To analyze determinants of the skill premium, the macroeconomic literature often uses production functions with technology-skill complementarity that exhibit decreasing marginal returns in the labor inputs. Using a production function in this class, [Krusell et al. \(2000\)](#) distinguish three key channels through which the skill supply (measured by the number of college graduates) can affect the skill premium (measured by the college premium). The first is the relative quantity effect given which a growth of skilled labor faster than that of unskilled labor reduces the skill premium. The second is the relative efficiency effect given which a growth of skilled labor efficiency faster than that of unskilled labor efficiency increases the skill premium. The third is the technology-skill complementarity effect.²⁰

As standard in the labor search literature, we assume a linear production in the efficiency units of labor, but we do have the aforementioned three channels in the model. First, the relative quantity effect operates not through decreasing marginal returns, but through frictional labor markets. In particular, when there is a relative increase in college graduates, the equilibrium contact rate from firms' perspective increases (equation 9) and thus the marginal return to efficiency units of labor decreases towards the level of the frictionless labor market, which results in a drop in the college premium. We also have the relative efficiency effect as a faster growth of skilled labor efficiency widens the gap between the efficiency units of labor of the skilled workers and those of unskilled workers. Finally, we assume a functional form for h_s such that the production admits technology-skill complementarity (see [Section 4.4](#)).

Endogenous Skill Demand and Supply It is well-documented in the empirical literature that both skill demand and supply play an important role in shaping wage inequality over time (e.g., [Goldin and Katz 2007](#)). Although thus essential for studying the source of inequality, few papers, however, embedded these equilibrium forces in a structural framework. In particular, standard labor models typically focus on skill demand and market frictions, taking skill supply as given (e.g., [Postel-Vinay and Robin 2002](#)), whereas macroeconomic models with endogenous skill supply typically use error-component models and take skill demand as given (e.g., [Huggett et al. 2011](#)).²¹

In our model, both skill demand and supply are endogenous, which enables us to investigate

¹⁹The unemployment duration is 15 and 16 weeks for college graduates and high school graduates, respectively, in 1980, and 21 and 20 weeks in 2005.

²⁰The technology-skill complementarity emerges from a capital-skill complementarity in their model. That is given by the fact that skilled labor is more complementary to equipment capital than unskilled labor. Formally, it can be expressed by the elasticity of substitution between the unskilled and equipment capital being higher than that between the skilled and equipment capital.

²¹Exceptions are the papers employing a general equilibrium framework of endogenous skill demand and supply with an aggregate production technology (e.g., [Heckman et al. 1998](#), [Abbott et al. 2016](#)). However, in these papers, workers with the same skills and schooling levels obtain the same wage.

the source of inequality in a rich environment. The equilibrium demand/supply interplay in the model is the following. Workers' education choices (i.e., skill supply) can react to any change in labor market conditions including skill demand summarized by unemployment value U^s , which can be seen later in the optimal cutoff decision (18). In turn, firms' hiring decisions v_s (i.e., skill demand) respond to a change in worker distribution Ψ that affects the expected profits equation (4) and thus problem (5). Finally, the skill demand and supply are tightly linked in the equilibrium contact rate equation (9).

In [Appendix B](#), we also discuss "sheepskin effect", wage bargaining, and risk preferences.

4 Equilibrium Analysis

In this section, we first characterize the stationary equilibrium almost in closed form, which gives us much tractability and numerous economic insights. The analytical result is key to deal with heterogeneity of both workers *and* firms. Next, we show that the model entails education sorting of workers and labor market sorting of firms which together imply assortative matching. We then provide our novel wage decomposition equations. Finally, we discuss how they are related to skill-biased technological change.

4.1 Analytical Characterization of Equilibrium

Unemployment Rates Worker flows determine the equilibrium unemployment rate of submarket s , u_s . In a stationary environment, the ins and outs of unemployment must be equal. The ins are given by $(1 - u_s)\delta_s + \mu$, which represents the case of workers that lose their jobs due to job destruction or complete/quit schooling. The outs are given by $(\lambda_s^U + \mu)u_s$, which represents the case of workers that find a job or die. The stationary condition for this reads

$$u_s = \frac{\delta_s + \mu}{\delta_s + \mu + \lambda_s^U}. \quad (10)$$

The unemployment rate increases with the job destruction and decreases with the job finding.

Wage Equations and Worker Distribution The optimal wage equations derived in [Cahuc et al. \(2006\)](#) can be easily extended to our segmented labor submarket structure with multi-worker firms. For an employed worker working at a p -firm with the best outside option $p' \leq p$,

the solution to the bargaining problem is given by²²

$$\phi^s(p', p) = p - (1 - \beta) \int_{p'}^p \frac{\rho + \delta_s + \lambda_s^E(1 - F^s(x))}{\rho + \delta_s + \lambda_s^E\beta(1 - F^s(x))} dx. \quad (11)$$

The fraction of workers employed by a p -firm is given by

$$l^s(p) = \frac{\delta_s + \mu + \lambda_s^E}{[\delta_s + \mu + \lambda_s^E(1 - F^s(p))]^2} f^s(p),$$

where f^s is the density of F^s .²³

Profit per Worker and Optimal Vacancies We first show that the profit per worker is linear in efficiency units of labor.

Proposition 1 *The profit per worker is linear in efficiency units of labor, i.e., $\pi^s(w, p, z) = \pi^s(w, p) \cdot h_s(z)$ where*

$$\pi^s(w, p) = (1 - \beta) \int_{g^s(w, p)}^p \frac{\rho}{\rho + \delta_s + \lambda_s^E\beta(1 - F^s(x))} dx. \quad (12)$$

Proof can be found in [Appendix C](#). It follows that, assuming $\pi_0 = 0$, we can write the value of an unfilled vacancy posted as follows:

$$\Pi^s(p) = \frac{\int_{z \in \mathcal{Z}} h_s(z) d\Psi^s(z)}{\rho + \frac{u_s}{\lambda_s^U} + \frac{1-u_s}{\lambda_s^E}} \left[\frac{u_s}{\lambda_s^U} \pi^s(\phi_0^s(p), p) + \frac{1-u_s}{\lambda_s^E} \int_{x \in \mathcal{P}} \pi^s(\phi^s(x, p), p) l^s(x) dx \right]. \quad (13)$$

Importantly, this equation shows that $\Pi^s(\cdot)$ does not depend on the entire distribution of worker's skill, which is a formidable object, but just on the conditional mean of efficient units of labor. This straightforwardly gives us the following proposition.

Proposition 2 *Suppose the vacancy creation cost is iso-elastic and given by $\kappa(v) = \chi \frac{v^{1+\frac{1}{\xi}}}{1+\frac{1}{\xi}}$.*

²²For an unemployed worker matched with a p -firm, the wage function is given by:

$$\phi_0^s(p) = p - (1 - \beta) \int_b^p \frac{\rho + \delta_s + \lambda_s^E(1 - F^s(x))}{\rho + \delta_s + \lambda_s^E\beta(1 - F^s(x))} dx.$$

Hence, we establish $\phi_0^s(p) = \phi^s(b, p)$ for any $p \in \mathcal{P}$.

²³We can define the mass of workers employed by a p -firm at a wage less than or equal to w :

$$G^s(w | p) = \left[\frac{\delta_s + \mu + \lambda_s^E(1 - F^s(p))}{\delta_s + \mu + \lambda_s^E[1 - F^s(g^s(w, p))]} \right]^2.$$

The unconditional distribution of accepted wages is then given by $\int_{p \in \mathcal{P}} G^s(w | p) l^s(p) dp$, i.e., the mass of workers that earn a wage less than or equal to w .

Then the optimal vacancy created by a p -firm is given by

$$v_s(p) = \left(\frac{\Pi^s(p)}{\chi} \right)^\xi. \quad (14)$$

Moreover, if the vacancy creation cost is quadratic, i.e., $\xi = 1$, then the optimal vacancy is linear in the conditional mean of efficient units of labor in submarket s .

This linear property of the optimal policy proves very useful in the computation of equilibrium, and hence we assume the quadratic vacancy creation cost in our quantitative analysis.

Value of being Unemployed The following proposition shows that the value of being unemployed is also linear in efficient units of labor, and it depends on the contact rate that prevails in each submarket. As discussed in [Section 3.4](#), the connection between these two equilibrium objects plays an important role in shaping the distribution of workers.

Proposition 3 *The value of being unemployed is linear in efficiency units of labor, i.e. $U^s(z) = U^s \cdot h_s(z)$, where*

$$U^s = b + \beta \int_{x \in \mathcal{P}} \frac{\lambda_s^U (1 - F^s(p))}{\rho + \delta_s + \lambda_s^E \beta (1 - F^s(p))} dp. \quad (15)$$

Proof can be found in [Appendix C](#). The elasticity of the unemployment value with respect to the contact rate is increasing in bargaining power β . This implies that the higher the workers' share in the match rent is, the more sensitive the unemployment value, hence the schooling decision, is to worker flows.

To elucidate the role of the rent share for the link between the unemployment value and the contact rate, suppose workers have no bargaining power, i.e., $\beta = 0$. Equation (15) states that U^s becomes independent of the contact rate, $U^s = b$ for all s . The schooling decision is then not affected by any change in the labor market conditions and solely driven by return to education, as shown below. This thought experiment proves the choice of β crucial qualitatively and quantitatively, although it is often chosen arbitrarily or set to target an aggregate variable such as aggregate labor share. Our calibration in [Section 5](#) reveals that the schooling decision is indeed affected by the labor market conditions.

Optimal policies in the Schooling Stage Equation (2) can be used to characterize the cutoff $\bar{q}^s$ that triggers dropout. At this cutoff value, the worker must be indifferent between continuing education or entering the labor market. Moreover, the marginal benefit of continuing education for one more instant must 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: for all $s \in \mathcal{S}$ and $z \in \mathcal{Z}$,

$$V^s(\bar{q}^s(z), z) = U^{s-1} \cdot h_{s-1}(z), \quad \text{and} \quad V_q^s(\bar{q}^s(z), z) = 0. \quad (16)$$

To make further progress, we make the following assumption.

Assumption 1 *The return to education is linear in efficiency units of labor, i.e., $h_s(z) = R_s \cdot h(z)$ where $R_s \in \mathbb{R}$ for $s \in \mathcal{S}_0$ with $R_0 = 1$ and $h : \mathcal{Z} \rightarrow \mathbb{R}_+$.*

We then establish the following lemma that states the value of schooling can be expressed as a linear function of efficiency units of labor.

Lemma 1 *Given Assumption 1, the value of schooling can be written as $V^s(q, z) = V^s(q) \cdot R_{s-1} \cdot h(z)$, where $V^s(\cdot)$ satisfies the corresponding version of equation (2): for $s \in \mathcal{S}$*

$$\rho V^s(q) = \rho(1-c)b - V_q^s(q)\dot{q} + \theta^s(q) \left[\frac{R_s}{R_{s-1}} V^{s+1}(\hat{q}(q)) - V^s(q) \right], \quad \text{where } V^{s+1}(\cdot) \equiv U^S. \quad (17)$$

Proof only requires to plug $V^s(q) \cdot R_{s-1} \cdot h(z)$ into equation (2). The next proposition uses this result to characterize the optimal cutoff beliefs.

Proposition 4 *Given Assumption 1, the optimal cutoff beliefs in the schooling stage, $\bar{q}^s$, are independent of z and characterized by*

$$\bar{q}^s = \frac{\rho[U^{s-1} - (1-c)b] - \theta_0^s[V^{s+1}(\hat{q}(\bar{q}^s)) - U^{s-1}]}{\Delta\theta^s \left[\frac{R_s}{R_{s-1}} V^{s+1}(\hat{q}(\bar{q}^s)) - U^{s-1} \right]} \quad \text{for } s \in \mathcal{S}. \quad (18)$$

Proof is obvious from equations (16-17) and hence omitted. A worker pursuing schooling level s drops out of education whenever her belief falls below $\bar{q}^s$. This result is very useful because it clarifies that cutoff rules are common across workers and do not depend on workers' own skill. Two workers with the same initial skill could end up with different educational attainment, which is a consequence of the informational structure of the model. We maintain Assumption 1 in the rest of the paper.

Stationary Distribution Given a vector of the optimal belief cutoffs, $\{\bar{q}^s\}_{s \in \mathcal{S}}$, we can solve for the workers' stationary distribution, Ψ , in closed form. This is so useful that it drastically reduces the computational burden. The expressions are derived in [Appendix D](#).

4.2 Two-Sided Sorting and Assortative Matching

Education Sorting of Workers As discussed in [Section 3.4](#), our model entails education sorting (i.e., high skilled workers tend to be highly educated) whose degree is governed by the correlation between ability a and skill z . Given the characterization of the optimal schooling

decision, we are now well-positioned to understand this sorting mechanism.

Consider two workers with the same ability but z_H and z_L where $z_H > z_L$. When the correlation is strictly positive, the z_H -worker will enter the schooling stage with a higher initial prior, q_H , than q_L that is the prior of the z_L -worker. The optimal cutoff belief characterized by equation (18) is independent of z (Proposition 4), so we have $q_H - \bar{q}^s > q_L - \bar{q}^s$. This means that the z_H -worker can wait longer for occurrence of the completion shock before the prior falling below the cutoff, and hence is more likely to complete the schooling level than the z_L -worker, even though their ability is the same. This argument therefore establishes the following proposition.

Proposition 5 (Education Sorting) *If the correlation between ability and skill is strictly positive, then the cumulative schooling attainment distribution of high skilled first-order stochastically dominates that of low skilled, i.e., $\Psi(s|z) \leq \Psi(s|z')$, where $\Psi(s|z) \equiv \sum_{t \leq s} \Psi(z, t) / \sum_{t \in \mathcal{S}_0} \Psi(z, t)$, for any $s \in \mathcal{S}_0$ and $z, z' \in \mathcal{Z}$ such that $z > z'$, with strict inequality for some s .*

Labor Market Sorting of Firms Using equation (13), we can decompose the value of an unfilled vacancy posted into two components, $\Pi^s(p) = \pi^s(p) \times \mathbb{E}h_s(z)$, where the first term is the value of a vacancy per efficiency units of labor and the second term is the average efficiency units of labor in submarket s . Given Proposition 2, this equation determines in which submarket a firm creates relatively more vacancies and thus is more active.

We already know that $\mathbb{E}h_s(z)$ is increasing in s , because of the return to education and education sorting. However, how $\pi^s(p)$ varies with s depends on the productivity p . For simplicity, suppose there are only a submarket for high school graduates and one for college graduates. Productive firms that can easily poach a worker from another firm have an incentive to post more vacancies in the college market than in the high school market as the average worker quality is better in the college market. In contrast, unproductive firms that can hardly keep a worker have an incentive to specialize in the high school market. This is because they fear the occurrence of their workers being poached by productive rivals whose likelihood is higher in the college market. Thus, productive firms are more active in the college market, whereas unproductive firms are more active in the high school market, which results in sorting of firms.

The intuition of this sorting mechanism is clear, but an analytical proof is not possible for the general case, as the optimal vacancy creation and the sampling distribution $F^s(\cdot)$ are determined simultaneously as a fixed point (see equation 8). We thus show this numerically in Section 6. In particular, in our baseline calibration, we establish the first-order stochastic dominance of the sampling distribution, i.e., $F^s(p) \leq F^{s'}(p)$, for any $p \in \mathcal{P}$ and $s, s' \in \mathcal{S}_0$ such that $s > s'$, with strict inequality for some p .

Two-Sided Sorting and Assortative Matching The two-sided sorting in our model –

education sorting of workers and labor market sorting of firms – together implies an assortative matching of high skilled workers to productive firms. This emerges naturally from the education choices and the segmented labor markets structure in the general equilibrium.

4.3 Wage Decomposition

Between-Group Inequality We denote by $w(z, s, p, p')$ the wage paid to a (z, s) -worker working at a p -firm whose best outside offer was made by a p' -firm. Using the results in [Section 4.1](#), we can write

$$w(z, s, p, p') = R_s \cdot h(z) \cdot \phi^s(p', p). \quad (19)$$

We call $h(z)$ and $R_s \cdot h(z)$ as ex-ante and ex-post worker productivity (i.e., efficiency units of labor), respectively. Note for the newly employed, we have $p' = b$.

The average wage paid in submarket $s \in \mathcal{S}_0$ is thus given by

$$\mathbb{E}[w(z, s, p, p') \mid s] = R_s \times \mathbb{E}_{\Psi^s}[h(z)] \times \mathbb{E}[\phi^s(p', p) \mid s], \quad (20)$$

where $\mathbb{E}_{\Psi^s}$ is the expectation operator taken over the conditional distribution Ψ^s . This equation makes clear that the average wage is the product of three key components specific to submarket s . The first component is return to education, which is productivity enhancement associated with higher education. The second is the conditional expectation of ex-ante worker productivity, which we call worker composition. This can vary across s due to education sorting. The third is the expected wage per efficiency unit of labor determined by firms' productivity, which we call firm composition.

It is immediate that these three components play key role in shaping the between-group wage inequality measured by the ratio of average wage in submarket s' relative to that in submarket s . The contribution of each component is captured in the following decomposition equation: for each $s \in \mathcal{S}$, the between-group inequality can be decomposed as

$$\ln \left\{ \frac{\mathbb{E}[w \mid s']}{\mathbb{E}[w \mid s]} \right\} = \underbrace{\ln R_{s'}}_{\text{Return to education}} + \underbrace{\ln \left\{ \frac{\mathbb{E}_{\Psi^{s'}}[h(z)]}{\mathbb{E}_{\Psi^s}[h(z)]} \right\}}_{\text{Worker composition}} + \underbrace{\ln \left\{ \frac{\mathbb{E}\phi^{s'}(p', p)}{\mathbb{E}\phi^s(p', p)} \right\}}_{\text{Firm composition}}. \quad (21)$$

This equation illustrates how different compensation for different education groups (e.g., college and high school graduates) is shaped by supply and demand of labor. On the supply side, college graduates have ex-post worker productivity higher than high school graduates on average, because of return to education *and* worker composition (education sorting). On the demand side, composition of firms active in each submarket is different due to firm sorting.

It is also clear that, if one abstracts from any component in equation (21) in its structural

model, the resulting decomposition would be potentially largely biased. For example, if one omits heterogeneity in innate skill z , the effect of education sorting would be attributed to return to education. Likewise, if one models education purely as a process of learning own skill, the estimate of worker composition effect would be biased upwards.

Within-Group Inequality Within-group wage inequality is measured by the variance of log wages conditional on submarket s . Denoting $\hat{x} = \ln x$, we use law of total variance to get

$$\text{Var}(\hat{w}|s) = \underbrace{\text{Var}[\hat{h}(z)|s]}_{\text{Person effect}} + \underbrace{\text{Var}\left\{\mathbb{E}[\hat{\phi}^s(p', p) | p]|s\right\}}_{\text{Firm effect}} + \underbrace{\mathbb{E}\left\{\text{Var}[\hat{\phi}^s(p', p) | p]|s\right\}}_{\text{Effect of market frictions}}, \quad (22)$$

where the labels of each component follows [Postel-Vinay and Robin \(2002\)](#). The first term called as person effect measures ex-ante productivity heterogeneity. The second called as firm effect measures wage variances across firms. The third called as effect of market frictions measures the within-firm wage dispersion not explained by person effect. This last component arises purely because similar workers in a firm have different histories of wage offers.

Between-/Within-Firm Inequality As discussed in [Section 2](#), an important observation about the inequality trend is related to between-/within-firm inequality. This decomposition is given by

$$\text{Var}(\hat{w}) = \underbrace{\text{Var}(\bar{w}_p)}_{\text{Between-firm inequality}} + \underbrace{\mathbb{E}[\text{Var}(\hat{w} - \bar{w}_p|p)]}_{\text{Within-firm inequality}}, \quad (23)$$

where $\bar{w}_p \equiv \mathbb{E}[\hat{w}(z, s, p, p') | p]$ is the average log wages paid by a p -firm. The unconditional log wage variance is decomposed into the between-firm inequality (variation in average log wages across firms) and the within-firm inequality (average of log wage variation within the firm).

Assortative Matching The between-firm inequality above can be further decomposed:²⁴

$$\text{Var}(\bar{w}_p) = \text{Var}\left[\mathbb{E}(\hat{h} | p)\right] + \text{Var}\left[\mathbb{E}(\hat{\phi} | p)\right] + \underbrace{2\text{Cov}\left[\mathbb{E}(\hat{h} | p), \mathbb{E}(\hat{\phi} | p)\right]}_{\text{Assortative matching}}. \quad (24)$$

That is, the between-firm inequality consists of the variation of average workers' productivity (workers' fixed effect), the variation of firms' wage policy (firms' fixed effect), and the joint variability of them. We call the last term as assortative matching, matching of productive workers to high-wage firms, that arises from the two-sided sorting discussed in [Section 4.2](#).

²⁴Note that the first term on the right hand side is not zero because the firm's vacancy creation policy is different across submarkets.

4.4 Skill-Biased Technological Change

A leading explanation for the wage inequality trends in the literature is the skill-biased technological change (SBTC). At the basic level, the SBTC can be interpreted as a technological change biased towards the skilled rather than the unskilled (Violante 2008). The revolution of information technology has been witnessed over the past thirty years, and such technology, for example, favors the skilled over the unskilled by augmenting its relative labor productivity in the production process.

To implement this idea, assume $h(z) = \exp(z)$ so that the production technology exhibits technology-skill complementarity. The elasticity of substitution between skill z and firms' technology p is then $1 + 1/z$, which is decreasing in skill.²⁵ That is, the technology is more complementary to high skills than low skills. This specification allows us to experiment with the SBTC in the model by changing the technology distribution Γ . If a firm becomes more technologically advanced, it wants more skilled workers.

The SBTC affects the wage inequality via not only the skill demand but also the skill supply. We numerically analyze this in the next sections.

5 Calibration

We calibrate our model to the pre-SBTC U.S. economy in 1980 and the post-SBTC economy in 2005 before the great recession.²⁶ The data used are described in Appendix A. All the empirical moments are five-year time-averaged to mitigate the cyclical effects (e.g., the 1980 value is the average of 1978-1982). For details of our numerical implementation, see Appendix E.

5.1 Pre-determined Parameters

Several parameters are identified without the model and assumed fixed over time. The continuous-time economy is discretized so that one time interval corresponds to a month. The interest rate r is set so that the annual rate is 6%. The birth/death rate μ is chosen to reflect 40 years of expected working life. These imply the annual effective discount rate of 0.92. The value of leisure b is normalized to be 1. We assume a unit measure of firms, $m = 1$.

We consider three schooling stages and thus three submarkets, $\mathcal{S}_0 = \{0, 1, 2\}$. Those with no completion, $s = 0$, correspond to high school graduates. Those who complete the first schooling level, $s = 1$, are college dropouts or some college, whereas those who complete all the schooling

²⁵The elasticity of substitution for a two-factor production function $f(p, z)$ is defined as $\frac{d \ln MRT(p, z)}{d \ln p/z}$ and the MRT is given by $-\frac{dz}{dp} = \frac{df(p, z)/dp}{df(p, z)/dz}$. Thus, the elasticity of substitution between p and z becomes $ES = \frac{d \ln(h(z)/ph'(z))}{d \ln p/z} = 1 + 1/z$.

²⁶Lee et al. (2017) also use a similar calibration strategy, assuming two steady states in 1980 and 2005.

Table 1: **Calibrated Parameters**

Parameter	Symbol	Value		Target
		1980	2005	
Return to education	R_s	(0.96, 1.11)	(1.02, 1.14)	Educational attainment
Separation hazard	δ_s	(.015, .010, .003)	(.010, .008, .003)	Unemployment rate
Vacancy creation cost	χ	.0022	.0029	Unemployment duration
Search technology	ζ	.30	.30	Job-to-job transition rate
Firms' Pareto parameter	γ	11.40	8.78	
Normal variance, skill dist.	σ_N	.07	.20	Wage inequality measures (see the text)
Exponential, skill dist.	α_a	(6.44, 6.43)	(3.41, 2.75)	
Bargaining power	β	.216	.271	

levels, $s = 2$, are college graduates. This classification is only for the sake of interpretation.²⁷

Workers forgo a half of the flow utility while in school, $c = 0.5$. The common completion hazard (θ_1, θ_2) is set so that the average duration of college education equals 5 years. This is broadly consistent with the college duration in the U.S.²⁸ We assume that a worker spend 1.5 years to become a college dropout, and 3.5 years to finish the college, on average, reflecting the fact that a number of college dropouts leave the college without a two-year certificate.

We assume the high/low types graduate faster/slower than the average, and set $(a_0, a_1) = (0.8, 1.2)$. This means that, conditional on no dropping out, the high/low types graduate the college in 4/6 years on average. There are equally many high and low types in the economy.²⁹

We assume a quadratic vacancy creation cost function, $\xi = 1$ (see [Proposition 2](#)).

The summary of predetermined parameters is found in [Appendix E](#). We conducted sensitivity analyses for these parameters and found that the results are materially unchanged.

5.2 Calibrated Parameters and Identification

We explain how to choose the empirical moments that identify the model parameters. The calibration of the remaining parameters takes two steps to reduce the computational burden. First, some parameters are exactly identified by the corresponding empirical moments. We then jointly calibrate the rest of the parameters by minimizing the RMSE. The calibrated parameters are summarized in [Table 1](#).

Returns to Education The educational attainment distribution identifies returns to educa-

²⁷For example, one can think of s as s -years of schooling. However, such a finer classification requires a sufficient sample size for each s .

²⁸Among 2007-08 first-time bachelor's degree recipients (including "stopouts"), the median (average) time between postsecondary enrollment and degree attainment is 52.0 (75.7) months. Source: National Center for Education Statistics, "Profile of 2007-08 First-Time Bachelor's Degree Recipients in 2009", Table 2.8.

²⁹In the data, 44.2% of students received the degree in 48 months or less. Source: Ibid.

tion R_s . To see this, consider a worker with skill z . The expected wage when he/she completes schooling level s is given by $R_s \cdot h(z) \cdot \mathbb{E}\phi^s$ (an individual version of equation 20). Thus, absent of any change in the labor markets conditions, return to education is the most relevant determinant for the education decision from worker's perspective.

We obtain the estimates of $(R_1, R_2) = (0.96, 1.11)$ for 1980 and $(1.02, 1.14)$ for 2005. The return to education is nonlinear and much of the return is associated with completing college ($R_2 > R_1 \simeq 1$). These higher returns to the latter years of college than to the early years are empirically supported in the literature (Card and Krueger 1992, Altonji 1993). Our estimates largely reflect the trends in the education attainment; the college enrollment rate has increased from 55.0% in 1980 to 64.9% in 2005 (see Table 2), which is reflected in higher return to college R_2 in 2005 than in 1980, while the graduation rate (i.e., the number of CL divided by the number of SC plus CL) has decreased from 60.0% to 55.9%, which is reflected in lower marginal return to degree R_2/R_1 .

The estimated return to college is broadly in line with the micro estimates (Card 1999). Using the employer-employee matched data in Ohio, Engbom and Moser (2017) find that college graduates earn 17.6% more in weekly wages than associate's degrees holders, which corresponds and is close to our estimates of R_2/R_1 (15.7% for 1980 and 12.2% for 2005).

XX the results of Heckman et al.(2006) in the dynamic model with the uncertainty about net earnings and both with and without nonlinearities of log-earnings in schooling vary between 20% and 26%.

Labor Market Parameters The unemployment rate for each education group identifies the separation hazard rate through equation (10). This gives $(\delta_0, \delta_1, \delta_2) = (.015, .010, .003)$ for 1980 and $(.010, .008, .003)$ for 2005, expressed in monthly frequency terms. The unemployment rates, and thus the separation hazard rates, are decreasing in the level of education.

The expected unemployment duration in submarket s can be calculated by solving

$$\int_0^\infty \mu \left(\int_0^t \lambda_s^U \tau e^{-\lambda_s^U \tau} d\tau \right) e^{-\mu t} dt = \frac{\lambda_s^U}{(\lambda_s^U + \mu)^2}. \quad (25)$$

Through this equation, the unemployment duration for each education group identifies the contact rate for the unemployed, λ_s^U .³⁰ Notice, however, that the contact rate is endogenous in our model and pinned down by equation (9). Because the education attainment distribution (denominator) is also obtained in the data, the unemployment duration thus effectively identifies the determinants of the firms' vacancy creation policies (numerator), in particular, the vacancy

³⁰Since equation (25) is quadratic, 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 $\mu = 0$.

creation cost parameter χ in equation (14).³¹ For the employed, we have $\lambda_s^E = (1 - \zeta)\lambda_s^U$ and set $\zeta = 0.30$ so that the monthly job-to-job transition rate is 0.02 (Shimer 2005).

Firms' Productivity, Skill Distribution, Bargaining Firms' productivity distribution is assumed to be Pareto, $\Gamma(p) = 1 - p^{-\gamma}$ with Pareto parameter γ .

We assume the skill distribution conditional on ability is Exponentially modified Gaussian:

$$z = z_N + z_E, \text{ where } z_N \sim N(\mu_N, \sigma_N^2) \text{ and } z_E \sim \exp(\alpha_a).$$

This specification has several advantages. First, the labor productivity distribution becomes Pareto log-normal, $h(z) = \exp(z) \sim PLN(\mu_N, \sigma_N^2, \alpha_a)$, that nests the standard productivity distributions used in the literature. In particular, it approaches to log-normal if $\alpha_a \rightarrow \infty$ and to Pareto if $\sigma_N^2 \rightarrow 0$. Second, we do not need to assume that the distributions for a_0 and a_1 are systematically different; instead, the two distributions differ only in the exponential parameter α_a . This naturally introduces a correlation between skill and ability. If $\alpha_{a_1} < \alpha_{a_0}$, as is the case in our estimates, then skill and ability are positively correlated and the skill distribution for high types has a thicker right tail than that for low types. We set μ_N to normalize the unconditional expected ex-ante productivity, $\mathbb{E}h(z) = 1$. Thus, we assume the average productivity of high school students does not change over time but the dispersion of productivity may.

We choose $\theta \equiv (\gamma, \sigma_N, \alpha_{a_0}, \alpha_{a_1}, \beta)$ to target the wage inequality measures: (i) the between-group inequality, (ii) the within-group inequality, and (iii) the between-/within-firm inequality decomposition (equation 23), for each point in time. Firms' productivity distribution (γ) and skill distribution (σ_N, α_a) are directly relevant for the wage inequality decomposition (Section 4.3). Also, as described in Section 4.1, the rent sharing rule (β) links the unemployment value and the contact rate, and thus affects wage differentials across submarkets. Thus, we solve

$$\min_{\theta \in \Theta} [\mathcal{M}(\theta) - \mathcal{M}_{\text{data}}]^T \mathbf{I}_{\|\mathcal{M}\|} [\mathcal{M}(\theta) - \mathcal{M}_{\text{data}}], \quad (26)$$

where $\mathcal{M}_{\text{data}}$ is a vector of the six empirical moments (see Table 2) and $\mathcal{M}(\theta)$ is a vector of the corresponding model moments given θ . We use the identity weighting matrix. Since the model is highly non-linear, we apply a global optimization algorithm using quasi-random numbers generated by Sobol sequence (see Appendix E).

The Pareto parameter γ of the firms' productivity distribution drops from 11.40 to 8.78, consistent with our SBTC notion. This implies that firms have a higher aggregate skill demand

³¹We could calibrate vacancy creation cost for each submarket. However, we do not opt for that, because the unemployment duration is quite similar across educational groups in the first place. For example, the duration is 16 weeks for high school graduates and 15 weeks for college graduates in 1980, likewise 20 weeks and 21 weeks in 2005.

in 2005 than in 1980 because of the technology-skill complementarity. The bargaining parameter β , that is assumed to be common across submarkets, slightly increases from .216 to .271. This estimate seems reasonable. For example, [Flinn and Mullins \(2015\)](#) use the aggregate labor share to estimate it and find .23.³² Finally, the workers' skill distribution parameters $(\sigma_N, \alpha_{a_0}, \alpha_{a_1})$ substantially changes from (.07, 6.44, 6.43) to (.20, 3.41, 2.75). That is, the ex-ante productivity distribution becomes not only more dispersed but also more skewed, as the 2005 distribution has a thicker right tail indicated by the lower Pareto parameters. This qualitative change is already documented in the literature. For example, XXLS variance of return to college increases by 64%. XXcitation

Identification While the tight linkage between the structural parameters in θ and the targeted wage inequality measures is clearly established in [Sections 4.1](#) and [4.3](#), demonstrating a formal analytical identification is difficult. We instead show the identification by a numerical simulation and provide an intuition based on that. FigureXX shows that

5.3 Model Performance

The summary of model performance is found in [Table 2](#).

Exactly Matched Moments As described in the previous section, some empirical moments are exactly matched by choosing the corresponding parameters. In particular, the model replicates the education attainment distribution, the unemployment rate for each education level, the average unemployment duration and the monthly job-to-job transition rate. The last moment is taken from [Shimer \(2005\)](#).³³

Targeted Moments The model matches the six wage inequality measures very well by choosing five parameters in θ . Specifically, the model can reproduce the increasing trend of not only the between-group inequality (measured by the ratio of average wage of CL/SC relative to that of HS) but also within-group inequality (measured by variance of log wages). Moreover, it replicates the contribution of the between-firm wage variation to the overall wage variation, $\text{Var}(\bar{w}_p) / \text{Var}(\hat{w})$ in equation (23). This data moment is taken and reconstructed from [Song et al. \(2015\)](#) by dropping the time-varying worker characteristics components that our model does not capture.³⁴

External Validation [To be completed] XX Corr(a,z) or Corr(a,s) or variance of initial q.

³²[Cahuc et al. \(2006\)](#) estimate the bargaining parameter for different occupation and industry using French data. The resulting estimates range from 0.00-0.98. If one excludes the most skilled occupation category (i.e., executives, managers, engineers), the upper bound drops to 0.33.

³³[Shimer \(2005\)](#) estimates the job-to-job transition using various data sets including CPS and argues that each estimate has pros and cons. We target the value that is roughly the average across the estimates (0.02).

³⁴The time-varying worker characteristics components play a very minor role in their decomposition. See their Table 2 and also equation 4.

Table 2: **Model Performance**

Moments	Data		Model	
	1980	2005	1980	2005
Exactly Matched				
Education attainment dist.	(.450, .224, .336)	(.351, .293, .356)	—	
Unemployment rate	(.055, .041, .018)	(.055, .043, .024)	—	
Ave. unemployment duration	3.46	4.54	—	
Monthly job-to-job transition*	0.02	0.02	—	
Targeted				
Between-group: SC vs HS	7.3%	21.8%	7.3%	21.8%
Between-group: CL vs HS	41.6%	107.5%	41.6%	107.5%
Winthin-group: HS	0.181	0.28	0.181	0.28
Winthin-group: SC	0.201	0.29	0.201	0.29
Winthin-group: CL	0.292	0.44	0.292	0.44
Between-firm contribution**	33.6%	37.4%	33.6%	37.4%
Untargeted				
	.222	.264		

* The data are taken from [Shimer \(2005\)](#).

** This is the contribution of the between-firm wage variation to the overall wage variation taken and reconstructed from [Song et al. \(2015\)](#) by dropping the time-varying worker characteristics components. The 1980/2005 value is based on their 1980-86/2001-07 decomposition.

Belzil and Jörgen Hansen (2002, ECTA) the corr between market ability and realized schooling is 0.28. STEM Taber (2001): Another line suggests that the rise of computers may have changed the workplace (Krueger, 1993; Autor, Katz and Krueger, 1999). It is not clear which skills taught in college should be complementary with computer use, but it seems quite feasible that the type of individual who chooses to attend college might be more likely to excel at computer work. For example individuals who enjoy learning may be more adept at adjusting to the new technology and may be more likely to attend college.

XX Skewness, uncond variance, median to min ratio from the entire sample. Hornstein et al. No need for low b to generate a substantial dispersion. Only need a pre-match investment and the segmented market setting. 50-10 versus 90-50. Autor RES median to mean

XX Sorting: First order stochastic dominance.

XX Assortative matching

XX E transition: the model will imply that EUE switch is more likely for HS, whereas EE switch is more likely for college. Haltiwanger

XX model generated data and run AKM

XX The empirical evidence shows that wages are positively correlated with profitability. For instance, using a matched panel to control for worker and firm heterogeneity, Abowd et al.

(1999) demonstrate that high-skill workers are paid more and that profitability is higher for firms with more skilled workers.

XX Move to Results Section

Hendricksy Todd Schoellmanz (2014 JME): changes in the composition of student abilities (skills in our case) by educational attainment to explain quantitatively the entire rise in the college wage premium while simultaneously making it easier to reconcile the current college wage premium with human capital theory (low completion).

XX LLS: The low wage gains are not surprising once we realize that marginal students do not go to college because they have realistic, low expectations of their returns to college in the first place. We note that our result is much smaller than the estimates in the empirical literature, e.g. Card (1999). The main reason for this discrepancy is that we explicitly account for the "selection on gains" that arises when students decide whether or not to graduate from colleges, based on signals of their post-graduation returns. Although much of the literature addresses positive selection into college enrollment following Heckman et al. (2006) and others, typically the next stage of selection - whether a marginal college student will graduate or not - is not explicitly addressed, likely biasing upward the returns-to-college estimates, e.g., Carneiro et al. (2011).

XX LLS: Finally, our model allows us to decompose the wage distribution into ex ante heterogeneity vs. ex post risk, which is the focus of Cunha et al. (2005) and Chen (2008). Our decomposition results are consistent with their finding that much of the wage dispersion among more educated workers is predictable from individual heterogeneity.

XX Stagnation of average wage, fall in the wage for low skills.

6 Quantitative Analysis

XX

We start by focusing on the decomposition of the college premium according to equation (21). In 1980, college graduates were making 40% more than high school graduates. About one third of this premium is accounted for by returns to education, a number that aligns well with typical estimates of Mincerian regressions. The model attributes the rest to firms' composition. Relative to that benchmark, the 2005 witnessed an increase in the importance of worker heterogeneity that originated almost entirely from workers' composition. In fact, the increase in the college premium can be almost entirely attributed to this factor. This suggests that the significance of worker fixed effects in wage equations must have increased over time.

- Relative to 1980, firms at the top and at the bottom offer higher wages and firms in the middle offer lower wages.

- Sorting interacts with the *climb-the-ladder* effect. The former creates incentives for posting more vacancies in the college submarket, where high productivity firms participate more heavily. This strengthens the climb the ladder effect for middle-productivity firms, which can afford to offer lower wages than before.
- This highlight the following property: the effect of a shift in the composition of firms has ambiguous effects on the wage gaps. In principle, average wages could go up or down depending on how competition among firms is affected. It is possible that college wages go down due to stronger climb the ladder effects in the relevant part of the support of the firm distribution, and that high school wages go up precisely because of the opposite reasons.

We decompose within-group inequality according to equation (22). In 1980, the share of dispersion coming from the demand side (firm and friction effects) was larger for college graduates than for high school graduates. In 2005, the figure is the opposite. Our decomposition suggests that this change is mostly due to an increase in the person effect for college graduates. That is, for college graduates there is a considerably larger amount of dispersion determined by the supply side of the market. All sources of dispersion increased over time for all educational groups, except for the friction effect in the case of college graduates. Therefore, in the case of high school workers, the increase in within-group dispersion can be almost equally explained by supply and demand forces, although there is a reduced relative importance of frictional factors. In contrast, in the case of college graduates, it stands out both the increase in the importance of the person effect and the drop in the importance of the friction effect.

Our model sheds light on why these changes took place. Relative to 1980, the calibration for 2005 displays more dispersion in firm’s productivity and more correlation between workers ability and productivity. This feature alone translates into a higher person effect. One could take this at face value and claim that the model just unveils ex-ante heterogeneity that is not modeled. An alternative interpretation, which we favor, is that education in 2005 is quite different from 1980. Students are evaluated more frequently and rigorously. In our model that would mean that the number of stages that takes to complete college is larger and as

Table 3: **Decomposition of College Premium**

	1980		2015	
	Level	%	Level	%
Heterogeneity	10.9	31.5	48.1	65.9
Returns to Education	10.9	31.4	13.2	18.1
Workers Composition	0.0	0.1	34.9	47.9
Firms Composition	23.7	68.5	24.9	34.1

students do so, worker’s productivity is more disperse and the correlation between productivity and ability increases. By considering just three stages, our calibration is basically truncating this process and capturing what the distribution would be at some point between enrollment and graduation. In other words, rather than attributing these changes to changes in ex-ante productivity, we think of them as the result of changes in the way education works.

Under this interpretation, we can say that just as it was the case for the wage premiums, educational sorting prior to labor market entry as an important force in shaping the increase in within-group inequality over the last 30 years.

6.1 Unconditional Variance Decomposition

We now use our model to analyze the unconditional variance of log wages. Table XX shows that our model is consistent with the evidence documented by Song et al, regarding the importance of between-firm dispersion in general and of assortative matching in particular in explaining the increase in the variance of log wages³⁵.

How much of the unconditional variance is explained by within-firm inequality? We address this question as follows. We calculate the unconditional variance of log wages and subtract from it, equation (23) in expectations. This fraction came down from 66.3% in 1980 to 53.2% in 2005. In other words, between-firm inequality has been more important in explaining the change in the unconditional variance. This suggests that between-firm measures may capture most of the increase in worker heterogeneity and assortative matching components from the

Table 4: **Decomposition of Wage Dispersion by Educational Attainment**

	1980		2015	
	Level	%	Level	%
High School	0.189	100.0	0.283	100.0
Person	0.029	15.6	0.083	29.4
Firm	0.058	30.9	0.102	35.9
Friction	0.101	53.5	0.098	34.7
College	0.293	100.0	0.437	100.0
Person	0.030	10.1	0.205	47.0
Firm	0.076	25.9	0.125	28.5
Friction	0.187	64.0	0.107	24.5

³⁵In our model, 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.

Table 5: **Decomposition Unconditional Variance**

	1980	γ unchanged	z unchanged	2015
Within Firm	66.3	68.1	53.2	62.3
Worker Heterogeneity	13.3	51.5	10.7	38.5
Residual	53.0	16.6	42.5	23.8
Between Firm	33.7	31.9	46.8	37.7
Segregation	0.1	1.0	0.1	0.6
Firm Heterogeneity	30.1	21.7	43.1	28.6
Assortative Matching	3.5	9.3	3.6	8.4

Table 6: **Decomposition Unconditional Variance**

	1980	γ unchanged	z unchanged	2015
Within Firm	0.163	0.214	0.161	0.262
Worker Heterogeneity	0.033	0.162	0.033	0.162
Residual	0.130	0.052	0.129	0.100
Between Firm	0.083	0.100	0.142	0.158
Segregation	0.000	0.003	0.000	0.003
Firm Heterogeneity	0.074	0.068	0.131	0.120
Assortative Matching	0.009	0.029	0.011	0.035
Total	0.246	0.314	0.303	0.421

unconditional variance decomposition.

Counterfactuals

Skill Biased Technical Change (SBTC)

If there were no SBTC, γ would stay the same level as in 1980.

- Unconditional log wage variance is lower without SBTC. This is consistent with the literature arguing that the SBTC is a driving force of the inequality trend
- However, this is mainly due to the drop in within-group inequality. The drop of the within-group inequality comes from lower firm effect (less variation in average pay across firms) and less friction. Thus the climb the ladder would be more frequent.
- Between-group inequality without SBTC would be lower than 2005 value but only slightly

7 Conclusions

[To be completed] 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

Table 7: **Had γ remained unchanged: Model Performance**

	1980	Counterfactual	2005
Wage gap: SC vs. HS	8.2 %	20.4 %	22.1 %
Wage gap: CL vs. HS	41.3 %	100.9 %	107.5 %
Wage dispersion: HS	0.19 %	0.19 %	0.28 %
Wage dispersion: SC	0.19 %	0.23 %	0.30 %
Wage dispersion: CL	0.29 %	0.33 %	0.44 %

Table 8: **Had γ remained unchanged: Wage gaps decomposition**

	1980		Counterfactual		2015	
	Level	%	Level	%	Level	%
Heterogeneity	10.9	31.5	49.9	71.5	48.1	65.9
Returns to Education	10.9	31.4	13.2	18.9	13.2	18.1
Workers Composition	0.0	0.1	36.7	52.6	34.9	47.9
Firms Composition	23.7	68.5	19.9	28.5	24.9	34.1

Table 9: **Had γ remained unchanged: Wage dispersion decomposition**

	1980		Counterfactual		2015	
	Level	%	Level	%	Level	%
High School	0.189	100.0	0.185	100.0	0.283	100.0
Person	0.029	15.6	0.083	44.9	0.083	29.4
Firm	0.058	30.9	0.054	29.1	0.102	35.9
Friction	0.101	53.5	0.048	26.0	0.098	34.7
Some College	0.194	100.0	0.225	100.0	0.302	100.0
Person	0.030	15.2	0.108	47.9	0.106	35.2
Firm	0.063	32.3	0.063	27.8	0.104	34.5
Friction	0.102	52.5	0.055	24.3	0.092	30.3
College	0.293	100.0	0.333	100.0	0.437	100.0
Person	0.030	10.1	0.210	63.1	0.205	47.0
Firm	0.076	25.9	0.070	21.1	0.125	28.5
Friction	0.187	64.0	0.053	15.9	0.107	24.5

Table 10: **Distribution of Educational Attainment**

	1980	γ unchanged	z unchanged	2015
High School	45.0	37.7	40.0	35.1
Some College	22.4	29.3	31.2	29.3
College	32.6	33.0	28.8	35.6

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 advice. 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

- Abbott, B., G. Gallipoli, C. Meghir, and G. Violante (2016). Education policy and intergenerational transfers in equilibrium. mimeo.
- Abowd, J. M., R. H. Creedy, and F. Kramarz (2002). 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). High Wage Workers and High Wage Firms. *Econometrica* 67(2), 251–334.
- Altonji, J. G. (1993). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics* 11(1), 48–83.
- Athreya, K. and J. Eberly (2016). Risk, the college premium, and aggregate human capital investment. mimeo.

- 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). 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). Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics. *American Economic Review* 104(6), 1551–96.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press.
- 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: Theory and evidence. *Econometrica* 74(2), 323–364.
- Card, D. (1999). The causal effect of education on earnings. In *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 30, pp. 1801–1863. Elsevier.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* 69(5), 1127–1160.
- 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.
- Card, D. and A. B. Krueger (1992). Does school quality matter? returns to education and the characteristics of public schools in the united states. *Journal of Political Economy* 100(1), 1–40.
- Eckstein, Z. and G. J. van den Berg (2007). Empirical labor search: A survey. *Journal of Econometrics* 136(2), 531–564.
- Engbom, N. and C. Moser (2017). Returns to education through access to higher-paying firms: Evidence from us matched employer-employee data. *American Economic Review* 107(5), 374–378.
- 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.
- Heckman, J., L. Lochner, and C. Taber (1998). Explaining rising wage inequality: Explanations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics* 1(1), 1–58.

- Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). *Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond*, Chapter 7, pp. 307–458. Elsevier.
- Hendricks, L., C. Herrington, and T. Schoellman (2017). The changing roles of family income and academic ability for us college attendance. mimeo.
- Hendricks, L. and O. Leukhina (2017). How risky is college investment? *Review of Economic Dynamics* 26, 140–163.
- Huggett, M., G. Ventura, and A. Yaron (2011). Sources of lifetime inequality. *The American Economic Review* 101(7), pp. 2923–2954.
- Keane, M. P. and K. I. Wolpin (1997). The career decisions of young men. *Journal of Political Economy* 105(3), 473–522.
- Kremer, M. and E. Maskin (1996). 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. Ríos-Rull, and G. L. Violante (2000). Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica* 68(5), 1029–1054.
- Lee, D. (2005). An estimable dynamic general equilibrium model of work, schooling, and occupational choice. *International Economic Review* 46(1), 1–34.
- Lee, S. Y. T., Y. Shin, and D. Lee (2017). 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.
- Manski, C. F. (1989). Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon. *Economics of Education Review* 8(4), 305–312.
- 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). Natural “Natural Experiments” in Economics. *Journal of Economic Literature* 38(4), 827–874.
- Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica* 50(1), 97–109.
- Shimer, R. (2005). The cyclicalities of hires, separations, and job-to-job transitions. *Federal Reserve Bank of St. Louis Review* 87(4), 493–508.

- Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2015). Firming up inequality. Working Paper 21199, National Bureau of Economic Research.
- Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics* 87(3), 355–374.
- Stange, K. M. (2012). An Empirical Investigation of the Option Value of College Enrollment. *American Economic Journal: Applied Economics* 4(1), 49–84.
- Stinebrickner, T. and R. Stinebrickner (2012). Learning about academic ability and the college dropout decision. *Journal of Labor Economics* 30(4), 707–748.
- Trachter, N. (2015). Stepping stone and option value in a model of postsecondary education. *Quantitative Economics* 6(1), 223–256.
- van den Berg, G. and G. Ridder (1998). An empirical equilibrium search model of the labor market. *Econometrica* 66(5), 1183–1222.
- 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 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 ³⁶. 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 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 as follows. 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.

B More Discussion on the Model

Sheepskin Effect We assume that the labor productivity does not change over time before the worker completes a schooling level. The resulting discrete jump in the average wage after the completion is known as a "sheepskin effect" whose existence is empirically supported in the literature. Moreover, in the previous version of the model, we assumed the productivity follows a Wiener process with heterogenous drifts conditional on ability, but the analysis was effectively the same and only added complexity in estimating the process.

Wage Bargaining In our model, identical workers working at the same firm may earn different wages. This wage differential, called wage dispersion, stems from the different histories of past offers these workers have received. The source of wage dispersion is the wage raise within

³⁶This 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 from 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.

the firm ("climb the ladder"), one of the central features of wage bargaining models pioneered by [Postel-Vinay and Robin \(2002\)](#). On the other hand, wage posting models à la [Burdett and Mortensen \(1998\)](#) may also generate wage dispersion as an outcome of mixed strategies by firms ([Mortensen 2003](#)). However, this is true only with a discrete firms' productivity distribution, and, with a continuous one, the dispersion vanishes. Thus, the wage bargaining is crucial for analyzing the source of wage inequality.

Risk Preferences We assume workers are risk neutral, which greatly simplifies the analysis. Since the within-group inequality has increased over the past 30 years, it might be argued that using risk averse preferences would be more appropriate when studying the education choice. In the data, within-group inequalities went up for all the education levels and the size of increase is comparable, i.e., the variance of log wages has increased across the board by about 50% (see [Figure 1](#)). We thus argue that the education decision would be similar even if we used risk averse preferences.³⁷

C Proofs

This section provides proofs of the propositions. Throughout, we suppress the schooling level s .

Proof of Proposition 1. Rearranging equation (3) and integrating by parts, we get:

$$(\rho + \delta + \lambda^E)\pi(w, p) = \rho(p - w) - \lambda^E \int_{g(w, p)}^p \pi_w(\phi(x, p), p) \phi_1(x, p) F(x) dx. \quad (27)$$

Differentiating both sides with respect to w and applying Leibniz's rule yields:

$$\pi_w(w, p) = -\frac{\rho}{\rho + \delta + \lambda^E [1 - F(g(w, p))]}.$$

where we have used $\phi_1(g(w, p), p)g_w(w, p) = 1$, by definition of $g(w, p)$. Plugging this expression back in equation (27) and noting that $\rho(p - w) = \rho \int_{g(w, p)}^p \phi_1(x, p) dx$, we have:

$$\pi(w, p) = \rho \int_{g(w, p)}^p \frac{\phi_1(x, p)}{\rho + \delta + \lambda^E (1 - F(x))} dx. \quad (28)$$

From equation (11), the derivative of the wage equation with respect to the first argument is:

$$\phi_1(p', p) = (1 - \beta) \frac{\rho + \delta + \lambda^E (1 - F(p'))}{\rho + \delta + \lambda^E \beta (1 - F(p'))}.$$

Plugging this into equation (28) yields equation (12). ■

Proof of Proposition 3. It is immediate to show that U and W are linear in $h(z)$. Using

³⁷We could extend our analysis to risk averse preferences. For example, for CRRA preferences, [Postel-Vinay and Robin \(2002\)](#) find that "the model thus naturally delivers a log-linear decomposition of wages clearly separating the effect of ability on one side and the effect of labor market history on the other" (p.2305).

the rent sharing condition of the bargaining game between worker and firms, we can write:

$$\rho W(p, p) = \rho p + \delta [U - W(p, p)] + \lambda^E \beta \int_p^{\bar{p}} [W(x, x) - W(p, p)] dF(x) \quad (29)$$

$$\rho U = \rho b + \lambda^U \beta \int_b^{\bar{p}} [W(x, x) - U] dF(x) \quad (30)$$

Subtracting one from the other and applying integration by parts.

$$\begin{aligned} \rho(\tilde{W}(p) - U) = \rho(p - b) &+ \delta [U - \tilde{W}(p)] + \lambda^E \beta \left[\tilde{W}(\bar{p}) - \tilde{W}(p) - \int_p^{\bar{p}} \tilde{W}'(x) F(x) dx \right] \\ &- \lambda^U \beta \left[\tilde{W}(\bar{p}) - U - \int_b^{\bar{p}} \tilde{W}'(x) F(x) dx \right], \end{aligned}$$

where we introduced the notation $\tilde{W}(p) \equiv W(p, p)$. Making $p = \bar{p}$, we obtain the expression for $\tilde{W}(\bar{p})$, which we can plug back and obtain:

$$\tilde{W}(p) = U + \frac{\rho}{\rho + \delta + \lambda^U} \left[\int_b^p \frac{\rho + \delta + \lambda^E \beta (1 - F(x)) + \lambda^U \beta F(x)}{\rho + \delta + \lambda^E \beta (1 - F(x))} dx - \int_p^{\bar{p}} \frac{(\lambda^U - \lambda^E)(1 - F(x))}{\rho + \delta + \lambda^E \beta (1 - F(x))} dx \right].$$

where we used the fact that $\tilde{W}'(p) = \rho / (\rho + \delta + \lambda^E \beta (1 - F(p)))$. Plugging this equation into equation (30), we obtain:

$$U = b + \frac{\lambda^U \beta}{\rho + \delta + \lambda^U} \int_b^{\bar{p}} \left[\int_b^y \left(1 + \frac{\lambda^U \beta F(x)}{\rho + \delta + \lambda^E \beta (1 - F(x))} \right) dx - \int_y^{\bar{p}} \frac{(\lambda^U - \lambda^E)(1 - F(x))}{\rho + \delta + \lambda^E \beta (1 - F(x))} dx \right] dF(y).$$

Finally, applying the integration by parts to the second term yields equation (15). ■

D Stationary Distribution

In this section, we characterize the stationary distribution of workers in closed form.

Consider a worker born with an initial belief q_0 calculated by Bayes' rule:

$$q_0(z) = \frac{\psi_{a_1}(z)}{\Pr\{a = a_1\} \psi_{a_1}(z) + [1 - \Pr\{a = a_1\}] \psi_{a_0}(z)},$$

where $\psi_a(\cdot)$ is the probability density of skills conditional on $a \in \mathcal{A}$. Define by $q(t)$ the belief at time t without any completion shock, thus $q(0) = q_0$. Solving the differential equation (1), we obtain

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

We define a function $t(q, q') = \{t' - t | q = q(t), q' = q(t')\}$ as the length of time for which the belief falls from q to q' . This can be solved as

$$t(q, q') = \frac{1}{\Delta\theta^s} \ln \left[\frac{q(1 - q')}{q'(1 - q)} \right].$$

Using this expression, we can solve for the mass of workers in each submarket that only depends on the optimal belief cutoffs, $\{\bar{q}^s\}_{s \in \mathcal{S}}$.

We focus on the case of $S = 2$ for illustration purpose. Given the following expressions for the density of workers, $\hat{\Psi}_a^s$, the (conditional) stationary distribution $\Psi^s(\cdot)$ for $s \in \mathcal{S}_0$ can be solved as

$$\Psi^s(z) = \frac{\hat{\Psi}^s(q_0(z))}{\sum_{s \in \mathcal{S}_0} \hat{\Psi}^s(q_0(z))}, \quad \text{where } \hat{\Psi}^s(q_0) \equiv \sum_{a \in \mathcal{A}} \Pr(a) \hat{\Psi}_a^s(q_0).$$

High School Graduates These workers do not experience any completion shock before $t(q_0, \bar{q}_1)$. Hence, the mass of high school graduates is given by

$$\hat{\Psi}_a^0(q_0) = \mu e^{-(\mu + \theta_a^1)t(q_0, \bar{q}_1)} \int_0^\infty e^{-\mu\tau} d\tau = e^{-(\mu + \theta_a^1)t(q_0, \bar{q}_1)}.$$

Some College Attendees There are two possibilities to become some college attendees. Namely, one either drops-out right after completing the first schooling level or spends some time in the second level before quitting the school. Hence, the mass of some college attendees is given by

$$\begin{aligned} \hat{\Psi}_a^1(q_0) &= \mu \int_0^{t(q_0, \bar{q})} \theta_a^1 e^{-(\mu + \theta_a^1)s} \left[e^{-(\mu + \theta_a^2)t(\hat{q}(q(s)), \bar{q}_2)} \int_0^\infty e^{\mu\tau} d\tau \right] ds + \mu \int_{t(q_0, \bar{q})}^{t(q_0, \bar{q}_1)} \theta_a^1 e^{-(\mu + \theta_a^1)s} \left[\int_0^\infty e^{\mu\tau} d\tau \right] ds \\ &= \theta_a^1 \left[\frac{\theta_0^1 (1 - q_0) \bar{q}_2}{\theta_1^1 (1 - \bar{q}_2) q_0} \right]^{\frac{\mu + \theta_a^2}{\Delta\theta^2}} \frac{1 - e^{-\left[\mu + \theta_a^1 - \frac{\Delta\theta_a^1}{\Delta\theta_a^2}(\mu + \theta_a^2)\right]t(q_0, \bar{q})}}{\mu + \theta_a^1 - \frac{\Delta\theta_a^1}{\Delta\theta_a^2}(\mu + \theta_a^2)} + \frac{\theta_a^1}{\mu + \theta_a^1} \left[e^{-(\mu + \theta_a^1)t(q_0, \bar{q})} - e^{-(\mu + \theta_a^1)t(q_0, \bar{q}_1)} \right], \end{aligned}$$

where $\tilde{q} = \max\{\hat{q}^{-1}(\bar{q}_2), \bar{q}_1\}$. Note that the set $\{\hat{q}(q) \mid q \in [q_0, \tilde{q}]\}$ contains all possible beliefs of agents who complete the first schooling level. Workers whose belief falls below $\bar{q}_2$ become some college.

College Graduates Workers who complete the first schooling level become college graduates at the rate θ_a^2 . Hence, the mass of college graduates is given by

$$\begin{aligned} \hat{\Psi}_a^2(q_0) &= \theta_a^2 \int_0^\infty e^{-\mu\tau} d\tau \left\{ \mu \int_0^{t(q_0, \bar{q})} \theta_a^1 e^{-(\mu + \theta_a^1)s} \left[\int_0^{t(\hat{q}(q(s)), \bar{q}_2)} e^{-(\mu + \theta_a^2)\tau} d\tau \right] ds \right\} \\ &= \frac{\theta_a^1 \theta_a^2}{(\mu + \theta_a^2)(\mu + \theta_a^1)} \left[1 - e^{-(\mu + \theta_a^1)t(q_0, \bar{q})} \right] - \frac{\theta_a^1 \theta_a^2 \left(\frac{\theta_0^1 (1 - q_0) \bar{q}_2}{\theta_1^1 (1 - \bar{q}_2) q_0} \right)^{\frac{\mu + \theta_a^2}{\Delta\theta^2}}}{(\mu + \theta_a^2)} \left[\frac{1 - e^{-\left(\mu + \theta_a^1 - \frac{\Delta\theta_a^1}{\Delta\theta_a^2}(\mu + \theta_a^2)\right)t(q_0, \bar{q})}}{\mu + \theta_a^1 - \frac{\Delta\theta_a^1}{\Delta\theta_a^2}(\mu + \theta_a^2)} \right]. \end{aligned}$$

We have verified these expressions using a discrete-time approximation to our continuous-time economy.

E Numerical Method

This section describes the numerical procedure to calculate the stationary equilibrium. It relies on the analytical results derived in [Section 4.1](#), which reduce the computational burden

Table 11: **Pre-determined Parameters**

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
Measure of firms	m	1
# of schooling levels	S	2
Completion hazard	(θ_1, θ_2)	Ave school duration: 5 yrs (=1.5+3.5 yrs)
Learning ability	(a_0, a_1)	(0.8, 1.2), Ave school duration: 4/6 yrs for high/low type
Marginal cost of schooling	c	0.5
Cost of vacancy creation	ξ	Quadratic vacancy creation cost, 1

considerably. Also we describe how to find a global optimum in the calibration. The summary of predetermined parameters is found in [Table 11](#).

Stationary Equilibrium We present the algorithm to compute a stationary equilibrium.

- Step 1: Calculate the belief cutoffs $\bar{q}^s$ that deliver the distribution of educational attainments observed in the data, using the expressions in [Appendix D](#).
- Step 2: Choose an initial guess for returns to education R .
- Step 3: Solve the labor market stage.
- Step 4: Calculate the value of being unemployed U^s , using equation (15).
- Step 5: Calculate the implicit returns to education R' that rationalize $\bar{q}^s$ in equation (18).
- Step 6: Stop if R' and R are close enough. Otherwise, set $R = R'$ and iterate Steps 3-5.

We have checked numerically that the system behaves as a contraction and thus the equilibrium seems unique.

Global Optimization Algorithm To ensure that our solution θ^* to problem (26) is a global optimum, we use the following algorithm.

- Step 1: Set $i = 1$. Choose an initial guess θ_i .
- Step 2: Apply the Nelder-Mead simplex method to solve problem (26) and find a solution $\tilde{\theta}_i$.
- Step 3: Construct a quasi-random parameter vector $\hat{\theta}_i \in \Theta$ generated by Sobol sequence.
- Step 4: Set the new initial guess by $\theta_i^{new} = f(i)\tilde{\theta}_i^* + (1 - f(i))\hat{\theta}_i$ where $\tilde{\theta}_i^*$ is the best solution obtained up until iteration i and $f : N \rightarrow [0, 1]$ is an increasing weighting function.
- Step 5: Set $i = i + 1$, $\theta_i = \theta_{i-1}^{new}$ and go back to Step 2.
- Step 6: Stop if converged.