Inequality, Productivity Dispersion, and Supply of Skills*

Manuel Macera †   Hitoshi Tsujiyama ‡

July 2020

Abstract

We study how an increase in firm productivity dispersion affects inequality across and within education groups (between- and within-group inequality) in the U.S. We develop a model in which skill supply is determined by college education choices and intergenerational linkages, and skill demand is characterized by firms' recruiting behavior. The model features worker and firm sorting that implies assortative matching of high-skilled workers with high-productivity firms. An increase in firm productivity dispersion changes skill demand. We find that the effect of this change on between-group inequality is attenuated by skill supply responses via endogenous education decisions but amplified by those via intergenerational linkages. The change in within-group inequality is mainly driven by the demand change, while skill supply responses play a limited role.

JEL-Codes: E24, I24, J23, J24, J31.

Keywords: Inequality, Productivity dispersion, Skills, Education, Intergenerational linkages, Frictional labor markets, Sorting.

---

*We thank Naoki Aizawa, Cristina Arellano, Antoine Camous, Nicola Fuchs-Schündeln, Jonathan Heathcote, Leo Kaas, Loukas Karabarbounis, John Kennan, Georgi Kocharkov, Chiara Lacava, Ilse Lindenlaub, Alexander Ludwig, Zachary Mahone, Giuseppe Moscarini, Andreas Mueller, Roberto Pinheiro, Fabrizio Perri, Ctirad Slavik, Naoki Takayama, Satoshi Tanaka and Gianluca Violante for useful comments and discussions.

†Universidad Torcuato Di Tella. Av. Figueroa Alcorta 7350, C1428 Ciudad de Buenos Aires, Argentina. mmacera@utdt.edu

‡Goethe University Frankfurt. Theodor-W.-Adorno-Platz 3, 60629 Frankfurt, Germany. Tel:+49-(0)69-798-33829. hitoshi.tsujiyama@hof.uni-frankfurt.de

1 Introduction

A striking feature of the trend in inequality in the U.S. in recent decades has been its increase both across and within education groups (between- and within-group inequality).[1] Over the same period, there has also been an increase in the dispersion of firm productivity.[2] This paper studies how the increase in firm productivity dispersion affects inequality across different groups of workers with different skill levels using a framework with endogenous demand and supply of skills.

Changes in firm productivity dispersion have both direct and indirect effects on inequality. The direct effects operate through marginal products; when firm productivity and worker skills are complements in production, higher productivity dispersion naturally translates into greater inequality. The indirect effects arise from endogenous responses in the supply and demand for skills. Specifically, given a change in the labor market environment via direct effects, workers optimally adjust schooling decisions and skill formation, whereas firms change recruiting behavior and wage policies. The responses in the supply and demand for skills also affect how workers are matched with firms in the presence of labor market frictions. In principle, the indirect effects can either amplify or attenuate the direct effect of rising firm productivity dispersion. Therefore, exploring its implications requires considering these effects in an equilibrium framework. For this purpose, we develop a tractable model with heterogeneous workers and firms that features endogenous supply and demand for skills. The key innovation is that we characterize how the supply of skills and schooling decisions respond to changes in the labor market environment when labor markets are frictional.

We now describe the model. The initial skill of a worker is determined by a reduced form of intergenerational linkages as a function of an exogenous innate skill (*nature*) and endogenous parental wages (*nurture*). The lifetime of a worker is then divided into two stages. In the schooling stage, workers make tertiary education decisions. Education is a process that not only enhances skill levels but also sorts workers into different labor markets according to their educational attainment. At the end of this stage, workers differ by skill level and schooling attainment. In the labor market stage,

[1]See, e.g., Goldin and Katz (2007).

[2]See, e.g., Dunne et al. (2004) and Brynjolfsson and Saunders (2010). Kehrig (2015) provides evidence that productivity dispersion increased in both durable and non-durable goods industries between the 1970s and 2010s in the U.S. Studies documenting the link between rising inequality and increasing dispersion in firm productivity include Faggio et al. (2010), Card et al. (2013), Dunne et al. (2004), Barth et al. (2016), and Song et al. (2019).

workers and firms, which also differ by productivity level, are matched randomly in labor markets segmented by educational attainment, and workers can search on the job. In this economy, the skill supply is determined by intergenerational linkages and college education decisions, whereas the skill demand is characterized by firms' recruiting policies.

The model features sorting of workers and firms in equilibrium. Since the education process allows for selection on skill, workers with high precollege skill tend to be highly educated, i.e., education sorting of workers. In addition, high-productivity firms recruit more intensively in submarkets with higher education where workers tend to be more skilled, whereas low-productivity firms focus on submarkets with lower education to avoid competing with more productive rivals, i.e., labor market sorting of firms. This two-sided sorting results in the assortative matching of high-skilled workers with high-productivity firms.

In this framework, when firm productivity dispersion increases, there are more high-productivity firms that pay more for high-skilled workers and recruit college workers more intensively—stronger firm sorting—which increases the college premium and wage dispersion. This demand-side effect on inequality can be attenuated or amplified by responses in the supply of skills and college education decisions. In particular, it is attenuated because the higher college premium encourages college attendance. On the other hand, it is amplified because the higher dispersion of wages induces a higher dispersion of precollege skills through intergenerational linkages, and thus, there are more high-skilled workers who tend to finish college—stronger worker sorting.

Using the IPUMS-CPS, we calibrate the model to the U.S. in 1970 and 2015. The main driver of increasing earnings inequality is an increase in the dispersion of firm productivity, the magnitude of which is comparable to our estimate using firm-level data from Compustat. The model matches nicely between- and within-group inequality measures, as well as the distribution of educational attainment and various labor market statistics. The calibration is also externally validated by matching various untargeted moments: distributions of firm size and output, Kelley's skewness of wages, trends in between- and within-firm inequality, and assortative matching (Song et al. 2019), the elasticity of college attendance and graduation rates with respect to financial aid (Castleman and Long 2016) and estimates of intergenerational mobility (Chetty et al. 2014). We also provide suggestive evidence of worker and firm sorting.

To understand the channels through which increasing dispersion in firm productivity affects inequality, we perform two counterfactual exercises. First, we take the 1970 economy and increase the dispersion of firm productivity to the level calibrated for 2015, holding fixed the skill and schooling attainment distributions at those in 1970. We find that even absent skill supply responses, the measures of between- and within-group inequality are almost as large as in 2015. Specifically, the increase in between-group inequality of college graduates relative to high school graduates (college premium) is 87% of the observed increase from 1970 to 2015, and the increase in within-group inequality is 93% of the observed increase for college graduates and 69% for high school graduates.

This finding does not necessarily mean that the response of the supply of skills has little impact on inequality. To see how skill supply interacts with the increase in firm productivity dispersion, in the second counterfactual exercise, we allow for endogenous skill supply responses through college education decisions but still hold fixed the precollege skill distribution at that in 1970. We find that, absent intergenerational linkages, the increase in between-group inequality is only modest, i.e., 52% of the observed increase, whereas the increase in within-group inequality is still large, i.e., 78% of the observed increase for college graduates and 79% for high school graduates.

These results suggest three key findings: (i) the effect of the increase in firm productivity dispersion for between-group inequality is *attenuated* by endogenous education decisions but *amplified* by intergenerational linkages, (ii) these two forces almost cancel one another out in equilibrium, and (iii) the change in within-group inequality is mainly driven by the response of the demand for skills; the skill supply plays a limited role.

To elucidate the economic forces behind these results, we decompose the measures of inequality in 1970, 2015 and the two counterfactual exercises. In the case of between-group inequality, we derive a novel equation that decomposes it into *returns to education*, differences in average workers' skills across education groups (*worker skill difference*), and differences in average firms' wages across education groups (*firm pay difference*). We find that returns to education increase only slightly from 1970 to 2015, and most of the observed increase in between-group inequality is accounted for by the changes in the worker skill difference and the firm pay difference, explaining 36% and 61% of it, respectively. In the first counterfactual exercise, the increase in between-group inequality is, by construction, due exclusively to a larger firm pay difference. This is simply the outcome of stronger firm sorting; there are more high-productivity firms that have high demand for skills and intensively recruit college workers, but since the supply of college workers is fixed, this results in stiffer competition in the college submarket and thus a larger firm pay difference across education groups. This force is mitigated in the second counterfactual exercise since earning a college degree becomes more attractive because of the higher college premium, and thus, the supply of college workers increases. Finally, when we also introduce intergenerational linkages, i.e., in the 2015 economy, precollege skill levels become more dispersed, and the worker skill difference across education groups increases because of stronger worker sorting.

For within-group inequality, following Postel-Vinay and Robin (2002), we decompose it into worker skill variation (*person effect*), wage policy variation across firms (*firm effect*), and wage policy variation within each firm (*friction effect*). We find that in 1970, within-group inequality is almost entirely explained by person effects for all education groups; workers are paid differently within an education group because their skill levels are different. However, the extent of worker skill variation changes little from 1970 to 2015, and most of the observed increase in within-group inequality is explained by an increase in the firm and friction effects. In 2015, they constitute 26% and 24% of within-group inequality for high school graduates and 24% and 16% for college graduates. In contrast to between-group inequality, this composition changes little in the two counterfactual exercises, irrespective of whether and how we control for skill supply responses. This suggests that the increase in within-group inequality is mainly driven by the increase in firm productivity dispersion and associated responses of skill demand, which induces larger firm and friction effects. Higher productivity dispersion increases not only the variation of wages across firms but also that of outside offers, resulting in a higher variation of wages within a firm.

Related Literature

The demand and supply of skills play a crucial role in shaping wage inequality over time (Katz and Murphy 1992, Goldin and Katz 2007), but few papers have embedded both in a structural framework with heterogenous workers and firms. In particular, standard labor search models typically focus on skill demand but take skill supply as given, while macroeconomic models with human capital investment typically use error-component models and thus take skill demand as given.[3] An important exception is

[3]Some papers employ a general equilibrium framework with an aggregate production technology (e.g., Heckman et al. 1998, Abbott et al. 2019). However, in these papers, there is no heterogeneity on the labor demand side, so workers with the same skill receive the same wage.

Flinn and Mullins (2015), who consider a labor search model with schooling choices. In their model, the distribution of initial skills is taken as given, and schooling involves no uncertainty, which implies that selection into college education is perfect. In contrast, in our model, intergenerational linkages make the initial skill distribution an endogenous object, and schooling choices generate skill distributions with overlapping supports across education groups.

We consider a standard labor search model with on-the-job search, applying the wage protocol developed in Cahuc et al. (2006) to a segmented labor market economy with multiworker firms. To model education choices, we follow the recent dynamic schooling choice literature that differentiates the college enrollment decision from the graduation decision (e.g., Stinebrickner and Stinebrickner 2012, Lee et al. 2017), and we contribute to this literature by distinguishing the continuation decision from graduation, embodying the conceptual framework of Manski (1989). Education is modeled as a process that not only enhances workers' skill but also sorts them into the labor market according to their educational attainment. Our framework thus embeds two views from the literature on the role of education in shaping the productivity distribution: education as a productivity-enhancing technology (Becker 1964) and education as a productivity-revealing signal (Spence 1973). Krueger and Ludwig (2016) and Abbott et al. (2019) also construct models with intergenerational transfers and educational choices, but they focus on understanding the effects of different policies in the aggregate.

The presence of assortative matching has been widely documented in recent literature (e.g., Lise et al. 2016, Hagedorn et al. 2017, Lopes de Melo 2018, Bagger and Lentz 2018, Song et al. 2019). On the theoretical side, assortative matching results from factors such as production complementarity between worker and job characteristics (Lise et al. 2016, Lopes de Melo 2018), heterogeneity in search technology (Bagger et al. 2014, Bagger and Lentz 2018) or firms' screening (Helpman et al. 2010). In our model, assortative matching occurs by a different mechanism: two-sided sorting, i.e., education sorting of workers and labor market sorting of firms. This mechanism is empirically supported; using employer-employee matched data, Engbom and Moser (2017) show that higher education degrees help sorting towards high-wage firms and that this sorting explains a substantial part of the returns to college.

Empirical studies typically decompose wage variation through the estimation of Mincerian wage functions that include worker and firm fixed effects (see e.g., Abowd et al. 1999, Card et al. 2013). Relative to this literature, wage decomposition based on structural models has the advantage of being able to address the potential biases of a reduced-form wage decomposition. For example, by considering the dynamics of worker mobility, Postel-Vinay and Robin (2002) find much less variation in worker fixed effects than do static error-component models that typically attribute historical wage differences to worker fixed effects. Our paper contributes to the decomposition literature based on structural models by explicitly modeling the selection of workers on skill, the selection of firms creating jobs for different education levels, and the systematic variation in labor market risk with education, each of which potentially cause systematic biases in reduced-form wage estimation.[4] Moreover, by conducting counterfactual exercises, we can investigate the economic forces behind the observed rise in inequality.

Our results are broadly consistent with the findings of several empirical studies. Keane and Wolpin (1997) and Huggett et al. (2011) find that differences in initial conditions are important factors in observed earnings inequality, although these differences are exogenous in their framework. Chen (2008) finds that the large gap in within-group inequality between high school and college graduates is mostly explained by worker heterogeneity. Our findings also indicate that the change in the worker skill difference is important for explaining the trends in both between-group inequality (Hendricks and Schoellman 2014) and within-group inequality (Taber 2001, Lemieux 2006). Finally, by explicitly considering employer heterogeneity, from which the previous papers abstract, we find that changes in firm characteristics are also crucial to explaining the trends in inequality. This echoes the results of recent papers that emphasize the role of firms in accounting for earnings inequality (Card et al. 2013, Song et al. 2019).

Our model implies that a larger gap in average workers' skill across education groups changes firms' recruiting behavior, inducing stronger firm sorting and thus a larger firm pay difference. This interaction between the worker skill difference and the firm pay difference across education groups is similar to the theoretical result of Acemoglu (1999) that a sufficiently large productivity gap between skilled and unskilled workers, potentially induced by skill-biased technological change, can cause a qualita-

[4]For instance, if education choices are correlated with unobserved skill, the college premium reflects a mix of true returns to education and selection on skill. In addition, if high-productivity firms create more jobs for workers with higher educational attainment, it is difficult to isolate firm characteristics from returns to education. Finally, it is difficult to cleanly separate out wage differences due to worker characteristics from labor market luck because important measures of labor market risk (e.g., the chance of becoming unemployed) vary with education.

tive change in the composition of the jobs that firms create. Our analysis also suggests that an exogenous skill demand shock is attenuated by the endogenous response of college education decisions and amplified by the presence of intergenerational linkages. This result complements Acemoglu (1998), who suggests that an exogenous increase in the supply of skilled workers triggers the development of a skill-biased production technology that increases between-group wage inequality.

2 Trends of Inequality and Productivity Dispersion

This section documents trends of between- and within-group inequality and those of productivity dispersion in the U.S. between 1970 and 2015.[5]

2.1 Trends of Between- and Within-Group Inequality

We use earnings data from the Current Population Survey (CPS), restricting attention to white males aged 25-55 who are working full time.[6] To calculate the inequality measures, we first estimate Mincerian regressions of the following form:

$$\log w_{i,t} = \alpha + \sum_{s} \beta_{s,t} D_{s,t} + \delta X_{i,t} + \epsilon_{i,t},$$

where $w_{i,t}$ represents the earnings of worker i at time t, $D_{s,t}$ are dummies for each educational level s, and $X_{i,t}$ are cubic controls for experience (Card 1999).

Between-Group Inequality. We measure between-group inequality as the difference in average log earnings between each educational group and high school graduates. To calculate the average earnings of each group, we use the output of the Mincerian regressions to control for the effect of experience on earnings. We let $\log w_{s,t} = \mathbb{E}\left[\log w_{i,t} - \delta X_{i,t} \mid s\right]$ denote average log earnings of educational group s and calculate between-group inequality simply as:

$$\log w_{s,t} - \log w_{\text{HS},t} = \beta_{s,t} - \beta_{\text{HS},t},$$

where HS stands for high school graduates, and s corresponds to either college graduates (CL) or individuals who have attended some college (SC).

[5] Appendix A contains a detailed data description.

[6] We restrict attention to a somewhat homogeneous group. We drop women from the sample because the educational attainment and labor force participation of women have changed dramatically in recent decades for various reasons, not only those on which we focus in this paper.

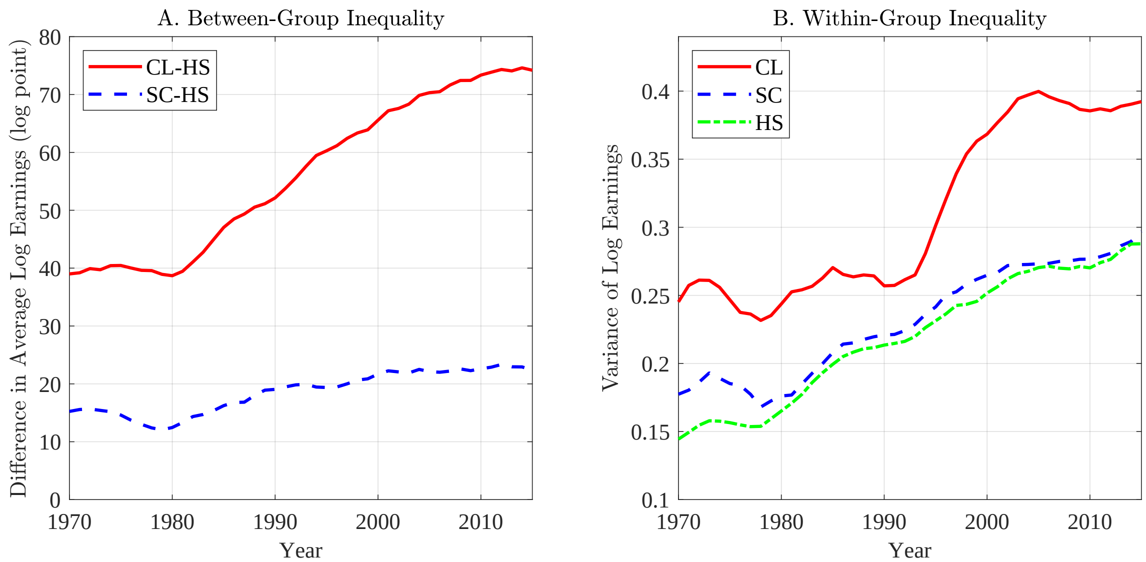

Figure 1: Between- and within-group earnings inequality. Panel A plots the between-group inequality measured as the difference in average log earnings between college graduates (CL) or some college (SC) attendees and high school graduates (HS) (log point). Panel B plots the within-group inequality of CL, SC and HS measured as the variance of log earnings. All values are five-year-centered moving averages.

Panel A of Figure 1 displays these measures. The college premium increased from 39 log points in 1970 to 74 log points in 2015. A similar trend is observed for the group of some college attendees, but the increase is much milder.

Within-Group Inequality. We measure within-group inequality as the conditional variance of log earnings for each educational group. Once we control for experience, this is simply the variance of the predicted residuals of the Mincerian regression:

$$\text{Var}\left[\log w_{i,t} - \log \hat{w}_{i,t} | s\right] = \text{Var}\left[e_{i,t} | s\right],$$

where $\hat{w}_{i,t}$ are predicted earnings and $e_{i,t}$ are predicted residuals.

Panel B of Figure 1 displays these measures. Within-group inequality increased over time for all categories. The variance of log earnings for HS and SC workers increased steadily over the past 45 years, from approximately 0.15 to 0.29, whereas for CL workers it remained flat at approximately 0.25 until the 1990s, then increased sharply to approximately 0.39, and became flat again in the mid-2000s.

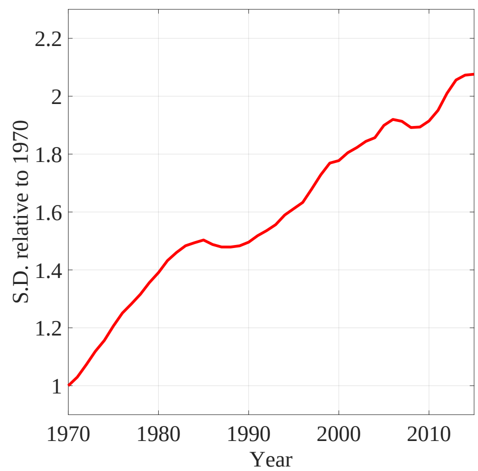

Figure 2: Productivity Dispersion. The figure plots the standard deviation of log output per worker relative to that in 1970. All values are five-year-centered moving-averaged.

2.2 Trends of Productivity Dispersion

We use firm-level data from Compustat, restricting attention to U.S. firms and excluding financial firms (SIC codes from 6000-6799). We measure firm productivity dispersion as the standard deviation of log output per worker (Decker et al. 2018). For calculating output, we use sales data because this information is widely available for most firms in Compustat, independently of their sector or size. Figure 2 plots this measure relative to that in 1970 and shows that dispersion has increased steadily over time; the standard deviation in 2015 is 2.1 times larger than that in 1970.[7]

As a cross-check, we also use value added to calculate output.[8] The correlation between value added per worker and sales per worker is high at 0.9. We find that this alternative measure of dispersion has also increased over time; the standard deviation of log value added per worker in 2015 is 2.6 times larger than in 1970 (Figure A1).

[7]Decker et al. (2018) also use revenue per worker as a measure of firm labor productivity and find that the productivity dispersion since mid-1990s has risen in the U.S.

[8]We do this only for robustness, because staff expenses, a necessary ingredient for value added, are only reported for 10.4% of firm-year observations. Following Hartman-Glaser et al. (2019), we estimate value added for firms that fail to report staff expenses by imputing per-worker labor cost with detailed industry-by-size means in each year. We use the Fama-French 17-industry classification and then sort firms into 20 size groups within each industry based on their total assets, yielding 340 industry-size cells in total.

3 The Model

Time is continuous and infinite. There is measure one of workers and measure one of firms. Workers face a constant birth/death rate μ, whereas firms live forever. Both types of agent are risk neutral and discount the future at a common discount rate r. We use $\rho = r + \mu$ to denote the worker's effective discount rate. The lifetime of a worker is divided into two stages: a schooling stage and a labor market stage. In the schooling stage, workers enhance their skill level prior to labor market entry. In the labor market stage, workers and firms are matched randomly, and production takes place.

3.1 Schooling Stage

In the schooling stage, workers make education choices. Education is a risky and costly process that enhances the workers' skills. Education is risky because it involves completing schooling levels, but completion is not guaranteed. Education is costly because workers incur an instantaneous marginal utility cost while in school.[9]

Each worker starts this stage with a schooling ability $a \in \{a_0, a_1\}$ with $a_0 < a_1$ and a skill level $z \in \mathcal{Z}$. We assume the type of a newborn depends on that of the worker from whom she descends, who is drawn at random from the living population. The schooling ability a of a newborn worker is correlated with parental ability, and the skill level z is given by

$$\log z = \log z_0 + \alpha_a \log(w), \tag{1}$$

where z_0 is an exogenous innate skill level and w is parental wage.[10] The parameter α_a measures the elasticity with respect to parental wage and depends on the newborn worker's learning ability. The idea behind this formulation is that parents influence the precollege skill levels of their children through both *nature* and *nurture*.[11] While this formulation buys much tractability, the intergenerational linkages we consider here is admittedly highly stylized. In Section 5, we empirically validate this specification by

[9]This utility cost represents all kinds of costs associated with education, including financial costs and psychic costs. Heckman et al. (2006) provide evidence on the importance of psychic costs in the schooling decision.

[10]Importantly, since the skill distribution affects the wage distribution, in equilibrium, the latter must be determined as a fixed point.

[11]This formulation captures the fact that resources invested in early childhood (proxied by wages) are important determinants of child development (e.g., Heckman 2006). It is also consistent with the evidence documented in Papageorge and Thom (forthcoming) that genetic traits interact strongly with economic status in determining educational outcomes.

showing that the intergenerational mobility implied by the model well matches with that estimated in Chetty et al. (2014).

There are S schooling levels indexed by s. The labor market productivity of a worker who completed level s becomes $R_s z$, which is measured in efficiency units of labor. The completion of each level occurs stochastically according to a Poisson process with arrival rate $\theta_i^s \equiv \theta_s a_i$. The parameter θ_s is related to the expected duration of schooling level s, whereas ability a_i captures the likelihood of completion. Workers who complete all schooling levels enter the labor market stage. Alternatively, they can choose to drop out and enter the labor market immediately, a decision that is assumed to be irreversible.

Worker's Problem. Workers know their skill but not their ability. Each holds a prior belief q of being a high-ability type. The individual state of a worker in the schooling stage is given by the triple (q, z, s), where s indicates the schooling level she is pursuing.[12] The prior belief is updated using Bayes' rule as described below. A worker with prior q pursuing schooling level s who experiences a completion shock will have a posterior $\hat{q}(q)$ given by:

$$\hat{q}(q) = \frac{\theta_1^s}{\theta^s(q)} q, \tag{2}$$

where $\theta^s(q) \equiv (1-q)\theta_0^s + q\theta_1^s$. In the absence of shocks, the prior belief will depreciate continuously according to[13]

$$\dot{q} = -\left(\theta_1^s - \theta_0^s\right)(1-q)q. \tag{3}$$

Workers solve an optimal stopping problem in which they must decide when to quit schooling. They take into account that the labor market is segmented by schooling attainment; a worker who quits schooling at level s will enter the submarket indexed by $s-1$.[14] The optimal decision takes the form of a cutoff belief $\overline{q}^s : \mathcal{Z} \to [0,1]$ that triggers dropout while pursuing schooling level s when $q = \overline{q}^s(z)$.

More formally, let V^s be the value of pursuing schooling level s while the option to drop out has not been exercised. This function must satisfy the following stochastic

[12] Hence, a worker starts this stage with state $(q(z), z, 1)$, where $q(z)$ is the subjective probability of being type a_1 given z, calculated using the cross-sectional joint distribution of skills and abilities.

[13] To see how belief depreciates, consider prior q_t at time t. The probability that the worker finishes schooling over the next interval dt is given by $\theta_a^s dt$. Thus, when she does not succeed, q_t becomes $q_{t+dt} = q_t(1-\theta_1^s dt)\left[q_t(1-\theta_1^s dt) + (1-q_t)(1-\theta_0^s dt)\right]^{-1}$. Subtracting q_t from both sides, dividing by dt and then taking $dt \to 0$ yields equation (3).

[14] To save on notation, we also use $s \in \{0, 1, \cdots, S\}$ to index schooling attainment.

partial differential equation:

$$\rho V^s(q,z) = \rho(1-c)b \cdot R_{s-1}z + \frac{\partial V^s(q,z)}{\partial q}\dot{q} + \theta^s(q)\left[V^{s+1}(\hat{q}(q),z) - V^s(q,z)\right], \quad (4)$$

where c denotes the constant marginal cost of schooling and b denotes the flow value of leisure. The first term on the right-hand side is the normalized instantaneous utility, which for tractability is assumed to be linear in efficiency units of labor. The second term captures the change in the value of schooling as a consequence of deterioration in the worker's belief. The last term represents the expected change in the value of schooling upon completion; in this case, the worker moves to the next schooling level $s+1$ with a new prior, $\hat{q}(q)$.

When the option to drop out is exercised, it must hold that $V^s(\overline{q}^s(z),z) = U^{s-1}(z)$, where $U^s(z)$ denotes the value of being unemployed in submarket s for a worker with skill z. Once in the labor market, schooling ability and, hence, the belief q become irrelevant. Finally, if $s = S$, upon completion, the worker enters the labor market with the highest educational attainment, and we thus have $V^{S+1}(\hat{q}(q),z) = U^S(z)$ in equation (4).

3.2 Labor Market Stage

There is heterogeneity on both sides of the labor market. The individual state of the worker consists of the pair (z,s). We use Ψ to denote the joint distribution of worker type and Ψ^s to denote the distribution of skills conditional on educational attainment. Firms are heterogeneous in their productivity level $p \in \mathcal{P} \equiv [b, \overline{p}]$, and Γ denotes the distribution of firms' type.

The labor market is segmented by the level of schooling attainment. In each submarket, workers and firms are matched randomly. Firms can post vacancies in all submarkets, but workers can only participate in the submarket corresponding to their educational attainment. Upon matching, worker and firm productivities are revealed to both parties. Jobs created in submarket s are destroyed at an exogenous rate δ_s.

Firm productivity and worker skills are complements in production; 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 $p \cdot R_s z$. The total output of a p-firm is equal to p times the sum of its employees' efficiency units of labor across all submarkets.

Wage Determination. A worker participating in submarket s will be contacted by a firm at rate λ_s^U if unemployed and λ_s^E if employed. We consider a different search technology for employed workers by assuming $\lambda_s^E = (1-\zeta)\lambda_s^U$ with $\zeta \in (0,1)$. The type of the firm is drawn from the sampling distribution F^s. Upon matching, the worker and the firm bargain over the wage under complete information.

We briefly describe our wage bargaining framework, which follows closely Cahuc et al. (2006). Without loss of generality, we assume that wages are set in terms of efficiency units of labor. Consider a (z, s)-worker who is contacted by a p-firm. If the worker is unemployed, bargaining results in a wage $\phi_0^s(p)$. If the worker is employed by a firm with productivity $p' < p$, the new firm poaches the worker from the incumbent, and bargaining results in a wage $\phi^s(p', p)$, where the first argument denotes the productivity of the last employer.[15] If the worker is employed by a firm with productivity $p' \geq p$, the incumbent can successfully deter poaching. In this case, however, bargaining resumes and results in a wage raise from the current wage w to $\phi^s(p, p')$ if and only if $p \in (g^s(w, p'), p']$, where $g^s(w, p')$ is the productivity level satisfying $\phi^s(g^s(w, p'), p') = w$.

Firms. Let $\pi^s(w, p, z)$ denote the profit for a p-firm hiring a (z, s)-worker at wage w. This profit must satisfy:

$$\begin{aligned} \rho\pi^s(w,p,z) &= \rho(p-w)\cdot R_s z - \left[\delta_s + \lambda_s^E(1-F^s(p))\right]\pi^s(w,p,z) \\ &\quad + \lambda_s^E \int_{g^s(w,p)}^{p} \left[\pi^s(\phi^s(x,p),p,z) - \pi^s(w,p,z)\right] dF^s(x). \end{aligned} \tag{5}$$

The first term on the right-hand side is the (normalized) flow profit. The second term is the expected capital loss that stems from either separation or poaching. The third term is the expected capital loss caused by the wage raise necessary to deter poaching.

To hire a worker with schooling attainment s, a firm must contact her in the corresponding submarket. Once a contact is made, there are three possible scenarios: the firm hires an unemployed worker, it hires a worker previously employed at a firm with lower productivity, or it fails to hire a worker because she is already employed at a firm with higher productivity. The expected profit per worker contacted $\Pi^s(p)$ thus

[15] In the class of labor search models with on-the-job search and wage renegotiation, wages typically depend on the individual history of past offers. In the present model, however, p' is a sufficient statistic for this history.

satisfies

$$r\Pi^s(p) = \frac{u_s}{u_s + (1-u_s)(1-\zeta)} \int_{z\in\mathcal{Z}} \pi^s(\phi_0^s(p), p, z) d\Psi^s(z)$$
$$+\frac{(1-u_s)(1-\zeta)}{u_s + (1-u_s)(1-\zeta)} \int_{z\in\mathcal{Z}} \left\{ \int_b^p \pi^s(\phi^s(x,p), p, z) l^s(x) dx \right\} d\Psi^s(z), \quad (6)$$

where u_s is the unemployment rate in submarket s, and $l^s(p)$ is the mass of workers employed by a generic p-firm. The first (second) term on the right side represents the expected profit, conditional on an unemployed (employed) worker being hired.

Since the production technology displays constant returns to scale, the recruiting decision is made separately for each submarket. Denote the contact frequency as ηv, where v is a measure of recruiting effort, call it "vacancies" as in Mortensen (2003), and η is the aggregate efficiency of recruiting effort.[16] For each s, a p-firm chooses a recruiting effort policy $v_s(p)$ to solve:

$$\max_{v_s} \left[\eta_s v_s \Pi^s(p) - \kappa(v_s)\right], \quad (7)$$

where $\kappa(\cdot)$ is a strictly increasing and strictly convex function that represents recruiting costs. We assume free entry.

Workers. Consider a (z, s)-worker. Let $U^s(z)$ be the value, to this worker, of being unemployed and $W^s(w, p, z)$ be the value of being employed by a p-firm at wage w. These values must jointly satisfy the following functional equations.

$$\rho W^s(w,p,z) = \rho w \cdot R_s z + \delta_s \left[U^s(z) - W^s(w,p,z)\right]$$
$$+\lambda_s^E \int_{g^s(w,p)}^p \left[W^s(\phi^s(x,p), p, z) - W^s(w,p,z)\right] dF^s(x)$$
$$+\lambda_s^E \int_p^{\overline{p}} \left[W^s(\phi^s(p,x), x, z) - W^s(w,p,z)\right] dF^s(x), \quad (8)$$

$$\rho U^s(z) = \rho b \cdot R_s z + \lambda_s^U \int_b^{\bar{p}} \max\left\{W(\phi_0^s(x), x, z) - U^s(z), 0\right\} dF^s(x), \quad (9)$$

The first term on the right-hand side of equation (8) is the wage paid. The second term represents the expected net loss due to exogenous separation. The third term is the expected net gain from a wage raise by the current employer, and the fourth is the

[16] Since v is a measure of recruiting effort made at a point in time, an unfilled "vacancy" is immediately destroyed.

expected net gain from accepting a wage offer made by a more productive employer. Finally, the second term on the right-hand side of equation (9) is the expected net gain from accepting an offer.

Sampling Distribution. The probability of being contacted with a p-firm in submarket s is proportional to the number of vacancies these firms create. Thus, the sampling distribution function is given by:

$$F^s(p) = \frac{\int_{x \leq p} v_s(x) d\Gamma(x)}{\int_{x \in \mathcal{P}} v_s(x) d\Gamma(x)}. \tag{10}$$

Contact Rates and Recruiting Efficiency. Each worker contacts firms at rate λ_s^U if unemployed and $(1-\zeta)\lambda_s^U$ if employed. Following Mortensen (2003), we assume that the aggregate flow of contacts per period in submarket s is proportional to the product of aggregate recruiting effort and workers' aggregate contact rate, i.e., $M_s = \lambda_s^U \eta_s$, where[17]

$$\lambda_s^U = \frac{\int_{x \in \mathcal{P}} v_s(x) d\Gamma}{\left[u_s + (1-u_s)(1-\zeta)\right] \int_{z \in \mathcal{Z}} d\Psi(z,s)}, \tag{11}$$

$$\eta_s = u_s + (1-u_s)(1-\zeta). \tag{12}$$

3.3 Stationary Equilibrium

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

(i) $\{\phi_0^s, \phi^s\}$ solve the wage bargaining problem;

(ii) $\{\overline{q}^s\}$ and $\{v_s\}$ solve the schooling decision and the recruiting decision problems;

(iii) $\{V^s, W^s, U^s\}$ and $\{\pi^s, \Pi^s\}$ satisfy their respective recursive equations;

(iv) $\{\lambda_s^U, \lambda_s^E, \eta_s, F^s\}$ are consistent with individual choices; and

[17] The aggregate flow of contacts is given by $M_s = \lambda_s^U \left[u_s + (1-u_s)(1-\zeta)\right]$. Analogously, because each p-firm contacts searchers at rate $\eta_s v_s(p)$, it also satisfies $M_s = m_s \eta_s \int_{x \in \mathcal{P}} v_s(x) d\Gamma(x)$, where m_s is the ratio of the measure of firms to that of workers searching a job in submarket s. It is given by

$$m_s = \frac{1}{\left[u_s + (1-u_s)(1-\zeta)\right] \int_{z \in \mathcal{Z}} d\Psi(z,s)}.$$

These expressions give equations (11-12).

(v) Ψ is stationary and consistent with individual choices.

Note that the intergenerational linkages connect the distribution of workers Ψ to the equilibrium wage distribution, as equation (1) makes clear. This means that the consistency of Ψ required in condition (v) adds an extra fixed point problem in the computation of equilibrium.

3.4 Discussion

In this section, we discuss some of the model assumptions and implications in detail.[18]

Model of Education. Education models based on Becker (1964) assume that individuals can choose the optimal years of schooling to maximize their expected utility (e.g., Keane and Wolpin 1997). 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, but they abstract from uncertainty in completion.[19] Following Manski (1989), we further distinguish continuation from graduation; in our model, pursuing a degree is discretionary, but obtaining that degree is uncertain.[20] This distinction is empirically relevant for two reasons.[21] First, survey data indicate that many individuals who plan to complete college drop out (Altonji 1993). Second, if those who fail to complete college are systematically different, in terms of labor productivity, from those who actually complete college, then abstracting from completion uncertainty results in a biased estimate of the returns to education.

We assume that skill z is observed but ability a is not. The former assumption is made only for simplicity; even if workers know only the expected value of z, the educational attainment decision remains unchanged. The latter assumption allows endogenous college dropouts. There are two advantages of this modeling choice. First, we can explicitly differentiate productivity enhancement from selection and, thus, returns to education from differences in worker composition as determinants of the measured college premium. Second, we can solve for the workers' stationary distribution in closed

[18]Further discussion can be found in Appendix B.

[19]See, e.g., Stange (2012), Stinebrickner and Stinebrickner (2012), Lee et al. (2017).

[20]The uncertainty in completion is different from the exogenous failure risk 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 the student's ability, which affects the probability of earning credits.

[21]It is also theoretically motivated by Manski (1989). He argues that for schooling to be completed, the student must not only persist until graduation but also be able to pass the courses. He states, "completion has both exogenous and endogenous determinants" (p.305).

form.

We allow ability a and skill z to be positively correlated. This means that the higher the observed skill level is, the higher the initial prior. Because the education decision problem boils down to an optimal stopping problem, schooling in our model entails that high-skilled workers also tend to be highly educated, i.e., education sorting.[22] 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 the returns to education. In contrast, if a and z are perfectly correlated and ability is thus observed, then the model would deliver perfect sorting, corresponding to the case of Flinn and Mullins (2015).[23]

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

Segmented Labor Markets. Labor markets are assumed to be segmented by schooling levels. This assumption aligns with signaling models à la Spence (1973) in which educational attainment partially reveals workers' productivity. Schooling also functions as a signal in our model because firms post vacancies knowing the average worker productivity in each submarket. Unlike in signaling models, however, in our model, education is not wasteful.

We rule out the possibility that college graduates take jobs that do not require college degrees (see, e.g., Lee et al. 2017), although the empirical relevance of this option is not conclusive in the literature. In the context of labor search, the presence of such an option implies a shorter unemployment duration for college graduates than for high school graduates since the former have a higher job-finding rate; however, the

[22] Manski (1989) calls the process of learning about academic ability "schooling as experimentation", although he does not explicitly consider learning in his model. Learning 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 their performance in school. Hendricks et al. (forthcoming) also document that academic ability is a more important predictor of who attended college after World War II than socioeconomic status.

[23] Since education is no longer a risky investment, only high-skilled workers finish college, and schooling splits the continuous support of initial skills into disconnected intervals. This 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.

observed durations are quite similar for these two groups.[24]

Determinants of College Premium. The macroeconomic literature often distinguishes three channels through which the supply and demand for skills can affect the college premium (see, e.g., Krusell et al. 2000). The first is the relative quantity effect in which faster growth in skilled labor, measured as the number of college graduates, relative to that of unskilled labor, measured as the number of high school graduates, reduces the skill premium. The second is the relative efficiency effect in which faster growth in the labor efficiency of college graduates relative to that in the labor efficiency of high school graduates increases the college premium. The third is the technology-skill complementarity effect.[25]

We also have these three channels in the model. First, the relative quantity effect operates through frictional labor markets. In particular, when there is a relative increase in college graduates, it makes relatively harder for them to be contacted by a firm, which pushes down their wages and results in a decrease in the college premium. We also have a relative efficiency effect, as faster growth of productivity of college graduates widens the gap between the average productivities of college and high school workers. Finally, firm productivity and worker skills are complements in production.

4 Equilibrium Analysis

In this section, we first characterize the stationary equilibrium in nearly closed form, which provides tractability and numerous economic insights. These analytical results are essential to address the computation of equilibrium when there is heterogeneity on both sides of the labor market. We then provide two decomposition equations for between- and within-group inequality. Finally, we argue that the model entails education sorting of workers and labor market sorting of firms, which together imply assortative matching.

4.1 Analytical Characterization of Equilibrium

Unemployment Rates. Worker flows determine the unemployment rate in each submarket. Let u_s denote the unemployment rate of submarket s. Inflows to unem-

[24]The unemployment duration was 13 weeks for both college and high school graduates in 1970, and 31 weeks for college graduates and 27 weeks for high school graduates in 2015.

[25]Krusell et al. (2000) consider a production function that exhibits decreasing marginal returns in the labor inputs. Technology-skill complementarity emerges from capital-skill complementarity in their model because skilled labor is more complementary to equipment capital than is unskilled labor.

ployment are given by $(1 - u_s)\delta_s + \mu$, which accounts for workers who either lose their jobs or complete/quit schooling. Outflows are given by $\left(\lambda_s^U + \mu\right) u_s$, which accounts for workers who either find a job or die. In a stationary equilibrium, inflows and outflows must be equal. Imposing this condition yields the following expression for the unemployment rate:[26]

$$u_s = \frac{\delta_s + \mu}{\delta_s + \mu + \lambda_s^U}. \tag{13}$$

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

Wage Equations and Worker Distribution. The optimal wage equations derived in Cahuc et al. (2006) can be easily applied to our segmented labor market structure with multiworker firms. For a worker employed at a p-firm with the best outside option $p' \leq p$, the solution to the bargaining problem is given by

$$\phi^s(p', p) = p - (1 - \beta) \int_{p'}^{p} \frac{\rho + \delta_s + \lambda_s^E (1 - F^s(x))}{\rho + \delta_s + \lambda_s^E \beta (1 - F^s(x))} dx, \tag{14}$$

where β denotes the worker's bargaining power.[27]

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

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

where f^s is the density of F^s.

Profit per Worker and Optimal Vacancies. We first show the following.

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

$$\pi^s(w, p) = (1 - \beta) \int_{g^s(w,p)}^{p} \frac{\rho}{\rho + \delta_s + \lambda_s^E \beta (1 - F^s(x))} dx. \tag{15}$$

The proof can be found in Appendix C. We can then write the expected profit per

[26] For each s, let E_s denote employed workers, U_s denote unemployed workers, and N_s denote new entrants. Inflows into unemployment are $\delta_s E_s + N_s$. Outflows are $(\lambda_s^U + \mu)U_s$. Stationarity requires not only inflows and outflows to be equal but also that new entrants make up for the workers that die, i.e., $N_s = \mu(E_s + U_s)$. Thus, in a stationary equilibrium, $(\lambda_s^U + \mu)U_s = \delta_s E_s + \mu(E_s + U_s)$. Using $u_s = U_s/(E_s + U_s)$ and rearranging yields equation (13).

[27] For an unemployed worker matched with a p-firm, the wage function is given by $\phi_0^s(p) = \phi^s(p_{\text{inf}}^s, p)$ for any $p \in \mathcal{P}$, where p_{inf}^s is defined below.

worker contacted as follows:

$$r\Pi^s(p) = R_s \frac{\int_{z\in\mathcal{Z}} z d\Psi^s(z)}{u_s + (1-u_s)(1-\zeta)} \left[u_s \pi^s(\phi_0^s(p), p) + (1-u_s)(1-\zeta) \int_{x\le p} \pi^s(\phi^s(x,p),p) l^s(x) dx \right]. \tag{16}$$

This equation makes clear that $\Pi^s(\cdot)$ does not depend on the entire distribution of worker skill but only on the average efficiency units of labor of workers participating in submarket s. This yields the following proposition.

Proposition 2 *Suppose that the recruiting cost is iso-elastic and given by* $\kappa(v) = \chi \frac{v^{1+1/\xi}}{1+1/\xi}$. *Then, the measure of recruiting effort made by a p-firm is given by*

$$v_s(p) = \left(\frac{\eta_s \Pi^s(p)}{\chi} \right)^{\xi}, \tag{17}$$

where the aggregate recruiting efficiency η_s *is given by equation (12).*

Notice that if the recruiting cost is quadratic, i.e., $\xi = 1$, then the optimal recruiting effort decision is linear in the conditional mean of efficiency units of labor in submarket s. This linear property of the optimal policy proves very useful in the computation of the equilibrium, and hence, we assume that $\xi = 1$ in our quantitative analysis.

It is worth noting that low-productivity firms might not be able to compensate an unemployed worker for the foregone value of unemployment and thus would not participate in a particular submarket. We denote by p^s_{inf} the productivity of the last firm that is able to attract a worker in submarket s.[28]

Value of Being Unemployed. We show that the value of being unemployed is also linear in efficiency units of labor and depends on the contact rate in each submarket.

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

$$U^s = b + \beta \int_{p^s_{\text{inf}}}^{\bar{p}} \frac{\lambda_s^U (1 - F^s(p))}{\rho + \delta_s + \lambda_s^E \beta (1 - F^s(p))} dp. \tag{18}$$

[28]The variable p^s_{inf} is defined as the productivity level that is just sufficient to compensate an unemployed worker for the forgone value of unemployment (Cahuc et al. 2006). The firm with p^s_{inf} pays the marginal product of labor and thus earn zero profits in submarket s. In equilibrium, this productivity value is given by

$$p^s_{\text{inf}} = b + \beta \zeta \lambda_s^U \int_{p^s_{\text{inf}}}^{\bar{p}} \frac{1 - F^s(x)}{\rho + \delta_s + \lambda_s^E \beta (1 - F^s(x))} dx.$$

The proof can be found in Appendix C. Since the sensitivity of the value of being unemployed with respect to the contact rate is increasing in bargaining power, the value of β is crucial both qualitatively and quantitatively. To elucidate its role in the schooling decision, suppose that workers have no bargaining power ($\beta = 0$). Equation (18) implies that U^s becomes independent of the contact rate. As shown below, this means that schooling decisions are not affected by changes in labor market conditions and are entirely driven by the returns to education. Our calibration in Section 5 requires $\beta > 0$ to match the data; thus, the schooling decision is indeed affected by labor market conditions.

Optimal Policies in the Schooling Stage. Equation (4) can be used to characterize the cutoff $\overline{q}^s$ that triggers dropout. At this cutoff value, the worker must be indifferent between continuing education and 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 captured by the following value matching and smooth pasting conditions:

$$V^s(\overline{q}^s(z), z) = U^{s-1} \cdot R_{s-1} z \quad \text{and} \quad \left.\frac{\partial V^s(q, z)}{\partial q}\right|_{q=\overline{q}^s(z)} = 0, \tag{19}$$

for all s and z. We then establish the following proposition.

Proposition 4 *The value of schooling is linear in efficiency units of labor, i.e., $V^s(q, z) = V^s(q) \cdot R_{s-1} z$, where $V^s(\cdot)$ satisfies:*

$$\rho V^s(q) = \rho(1-c)b - \frac{\partial V^s(q)}{\partial q}\dot{q} + \theta^s(q)\left[\tfrac{R_s}{R_{s-1}} V^{s+1}(\hat{q}(q)) - V^s(q)\right], \tag{20}$$

where $V^{S+1}(\cdot) \equiv U^S$.

The proof follows from plugging $V^s(q) \cdot R_{s-1} z$ into equation (4). The next proposition uses this result to characterize the optimal cutoff beliefs.

Proposition 5 *The optimal cutoff belief in the schooling stage is independent of z and characterized by*

$$\overline{q}^s = \frac{\rho\left[1 - (1-c)\frac{b}{U^{s-1}}\right] - \theta_0^s\left[\frac{R_s}{R_{s-1}}\frac{V^{s+1}(\hat{q}(\overline{q}^s))}{U^{s-1}} - 1\right]}{(\theta_1^s - \theta_0^s)\left[\frac{R_s}{R_{s-1}}\frac{V^{s+1}(\hat{q}(\overline{q}^s))}{U^{s-1}} - 1\right]}, \tag{21}$$

for all $s \in \{1, \cdots, S\}$, and $V^{S+1}(\cdot) \equiv U^S$.

The proof is straightforward from equations (19-20) and hence omitted. A worker pursuing schooling level s drops out of education whenever her belief falls below $\overline{q}^s$. Proposition 5 is very useful because it means that we do not have to solve for $\overline{q}^s$ as a function of z, which facilitates computation.

Equation (21) shows that the optimal cutoff beliefs hinge on three elements: (i) the marginal returns to education R_s/R_{s-1}, (ii) the marginal gain in the value of unemployment/schooling V^{s+1}/U^{s-1}, and (iii) the opportunity cost of schooling $U^{s-1}/(1-c)b$. The optimal cutoff decreases in R_s and V^{s+1} and increases in U^{s-1}. That is, schooling becomes more attractive, and workers thus stay longer in school, when there is an increase in the returns to education or in the value of pursuing a higher schooling level or when the value of dropping out and entering the labor market decreases. The latter is the case because the marginal gain from schooling increases, and the opportunity cost of staying in school decreases. Note that when the values of two consecutive schooling levels increase, the effect on $\overline{q}^s$ is ambiguous.

Stationary Distribution. Given the optimal belief cutoffs, we can solve for the workers' stationary distribution Ψ in closed form (see Appendix D for the expressions). This feature drastically reduces the computational burden. Importantly, Ψ^s has overlapping supports in z because the completion of schooling is stochastic.

4.2 Decomposition

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

$$w(z, s, p, p') = R_s z \cdot \phi^s(p', p). \tag{22}$$

For the newly employed, we have $p' = b$. The average log wage paid in submarket s is then given by

$$\mathbb{E}[\log w(z, s, p, p') \mid s] = \log R_s + \mathbb{E}[\log z|s] + \mathbb{E}[\log \phi^s(p', p)|s]. \tag{23}$$

Using this equation, we measure between-group inequality as the difference in average log wages between any two submarkets s and s', as in Section 2. Our first decomposition equation is:

$$\underbrace{\mathbb{E}[\log w \mid s'] - \mathbb{E}[\log w \mid s]}_{\text{BG inequality}} = \underbrace{\log R_{s'} - \log R_s}_{\text{Returns to education}} + \underbrace{\mathbb{E}[\log z|s'] - \mathbb{E}[\log z|s]}_{\text{Worker skill difference}} + \underbrace{\mathbb{E}[\log \phi^s(p', p)|s'] - \mathbb{E}[\log \phi^s(p', p)|s]}_{\text{Firm pay difference}}. \tag{24}$$

This equation makes clear that three components play a key role in shaping between-group wage inequality. The first component represents the difference in the pure *returns to education*. The second is the difference in the conditional mean of precollege log skills, which we label *worker skill difference*. The third is the difference in the conditional mean of log wages measured in efficiency units, which we label *firm pay difference*.

This equation also clarifies how between-group inequality is shaped by the supply of and demand for labor. On the supply side, average worker productivity can be higher for college graduates than for high school graduates not only because workers are different (worker skill difference) but also because they make different education choices (returns to education). On the demand side, the average wage in efficiency units of labor can be higher for college graduates than for high school graduates because the vacancy creation decision depends in a nontrivial way on firm characteristics (firm pay difference).

It is also clear that if a component in equation (24) were omitted from a structural model, the resulting decomposition would be biased. For example, if one abstracts from heterogeneity in skill, a worker skill difference is attributed to returns to education. Similarly, if education is modeled purely as a process of learning unknown types, the estimate of worker skill difference is biased upwards.

Within-Group Inequality. We measure within-group wage inequality as the conditional variance of log wages. By the law of total variance, we obtain our second decomposition equation:

$$\underbrace{\text{Var}\left[\log w|s\right]}_{\text{WG inequality}} = \underbrace{\text{Var}[\log z|s]}_{\text{Person effect}} + \underbrace{\text{Var}\left[\mathbb{E}[\log \phi^s(p', p) \mid p]|s\right]}_{\text{Firm effect}} + \underbrace{\mathbb{E}\left[\text{Var}[\log \phi^s(p', p) \mid p]|s\right]}_{\text{Friction effect}}, \tag{25}$$

where the labels of each component on the right-hand side follow Postel-Vinay and Robin (2002). The person effect measures ex ante skill heterogeneity. The firm effect measures variation in average log wages across firms. The friction effect measures average within-firm log wage variation, independent of the person effect. The latter effect arises purely because two identical workers in the same firm may have different wage offer histories.

4.3 Two-Sided Sorting and Assortative Matching

Education Sorting of Workers. Given the characterization of the optimal schooling decision, we are now well positioned to understand the mechanism of education sorting (i.e., high-skilled workers tend to be highly educated). We first introduce a cardinal measure of sorting.[29]

Definition 1 *The equilibrium displays positive education sorting if* $corr(s, z) > 0$ *among workers participating in the labor market.*

First-order stochastic dominance in the distribution of skills conditional on schooling attainment is sufficient but not necessary for positive sorting. In fact, the degree of sorting is governed by the correlation between ability a and skill z. To understand this, note that because the cutoff belief $\overline{q}$ is independent of z (Proposition 5), workers with higher initial prior q will wait longer for a completion shock and, hence, are more likely to complete any schooling level. This implies a positive correlation between the initial prior q and realized schooling attainment s. Since the initial prior is calculated using Bayes' rule, a positive correlation between a and z implies a positive correlation between q and z and, as a result, positive education sorting. Thus, in an equilibrium with positive education sorting, workers with higher precollege skill will have, on average, higher educational attainment.

Stronger (weaker) education sorting increases (decreases) unconditional earnings inequality. However, it has an ambiguous effect on the worker skill difference in equation (24) because it affects the average skill of the two education groups at the same time. Nevertheless, this sorting mechanism is still very helpful for understanding the sources of the trend in inequality, as we will show in Section 6.

Labor Market Sorting of Firms. We introduce an analogous definition for firms using the distribution of vacancies rather than employment because only the former affect workers' schooling decisions.

[29] In Section 5, we compute the Spearman rank correlation rather than the Pearson correlation as a measure of sorting that does not depend on the scale of z, thus allowing us to compare its degree over time. For the same reason, we use the rank correlation for the degree of firm sorting defined below.

Definition 2 *The equilibrium displays positive firm sorting if $corr(s, p) > 0$ among vacancies posted in the labor market.*

Again, first-order stochastic dominance in the distribution of firm productivity conditional on labor submarket is sufficient for positive sorting of firms.

We argue that our economy exhibits positive firm sorting—high-productivity firms tend to recruit more and are thus more active in submarkets with higher education. To understand this, we use equation (16) to write the expected profit per worker contacted as $\Pi^s(p) = r^{-1}\mathbb{E}\left[\pi^s(w, p) \mid p, s\right] \times \mathbb{E}[R_s z \mid s]$. The first term is the expected profit per worker expressed in efficiency units of labor, and the second term is the average efficiency units of labor in submarket s. According to Proposition 2, this equation determines in which submarket a firm creates more vacancies and is thus more active. Since the second term is independent of p, the degree of firm sorting depends on how $\mathbb{E}\left[\pi^s(w, p) \mid p, s\right]$ varies with p and s. The expected profit per worker is decreasing in the likelihood that prospective employees are contacted by a firm that would either poach the worker away or trigger a wage increase. High-productivity firms find it relatively easy to hire and retain employees since there are few competitors. They thus post more vacancies in submarkets with more productive workers, i.e., those with higher educational attainment. In contrast, low-productivity firms have difficulty retaining workers in submarkets in which high-productivity firms are more active. Therefore, they participate more actively in submarkets with lower educational attainment, where the likelihood that their employees are poached away is low. As a result, positive firm sorting occurs in equilibrium.

Two-Sided Sorting and Assortative Matching. Two-sided sorting—education sorting of workers and labor market sorting of firms—implies assortative matching of high-skilled workers with high-productivity firms, i.e., a positive correlation between z and p across matches. This feature of the equilibrium stems from the optimal education choices and the segmented labor market structure.

Note that assortative matching arises only if there are frictions in the labor market. With frictionless labor markets, all workers would work for the most productive firm, and measured assortative matching would thus be zero.

5 Calibration

We calibrate our model to the U.S. economy in 1970 and 2015.[30]

5.1 Predetermined Parameters

The parameters discussed in this section are fixed over time. We conducted sensitivity analyses, and the results were materially unchanged.

The continuous-time economy is discretized such that one time interval corresponds to one month. The interest rate r is set such that the annual rate is 6%. The birth/death rate μ reflects 40 years of expected working life. These parameters imply an annual effective discount rate of 8.5%. The value of leisure b is normalized to 1. We assume a quadratic vacancy creation cost function, $\xi = 1$.

We consider two schooling levels ($S = 2$) and, thus, three labor submarkets. These submarkets correspond to high school graduates (HS), some college attendees (SC), and college graduates (CL). This classification allows us to exploit labor market data by educational attainment.[31] Workers forgo half of the flow utility while in school, $c = 0.5$.[32] We assume that the average duration of college education is 5 years, which is broadly consistent with college duration in the U.S.[33] The common completion hazard (θ_1, θ_2) is then set such that a worker spends 1/3 of this duration to become SC and 2/3 to become CL, on average, reflecting the fact that a number of college dropouts leave without a two-year certificate. We assume the high (low) types graduate faster (slower) than average, and we set $(a_0, a_1) = (0.8, 1.2)$. This means that conditional on not dropping out, high (low) types graduate college in 4 years (6 years), on average. The fraction of high types in the population is set at 50%.[34]

[30] The details of the numerical procedure to calculate stationary equilibria are presented in Appendix E.

[31] One can think of s as indicating years of schooling. A finer classification, however, would require a sufficiently large sample size for each s.

[32] The costs of college are given by $c \cdot b \cdot \mathbb{E}h_s(z)$. In equilibrium, the implied average annual costs including financial and psychic costs are \$24,054 in 1970 and \$31,710 in 2015. Abbott et al. (2019) use data for 2000-2001 and estimate the average annual net financial costs of college at \$9,236 and the average total psychic costs of college at \$85,862. Assuming the average duration of college is 5 years, the average annual costs of college become \$26,408, which is in the middle of our model-implied costs of college for two points in time. All amounts are in 2015 dollars.

[33] Among 2007-08 first-time bachelor's degree recipients (including "stopouts"), the median (average) time between postsecondary enrollment and degree attainment was 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.

[34] In the data, 44.2% of students received a college degree in 48 months or less. Source: Ibid.

5.2 Calibrated Parameters

For the calibration of the remaining parameters, we proceed in two steps. First, some parameters are exactly identified by the corresponding empirical moments. We then jointly calibrate the remaining parameters using the method of simulated moments. The calibrated parameters and the targeted moments are summarized in Table A1.

Returns to Education. The distribution of educational attainment identifies the returns to education R_s. The expected wage when a worker with skill z completes schooling level s is given by $R_s z \cdot \mathbb{E}\phi^s$ (see equation 22), which indicates that returns to education are an important determinant of the education decision from the worker's perspective.

We obtain $(R_1, R_2) = (1.03, 1.20)$ for 1970 and $(1.03, 1.21)$ for 2015, which is broadly in line with micro estimates.[35] The calibrated pure returns to education increase slightly over time. These estimates also suggest that the returns are higher in the late rather than the early years of college and that most of these returns are associated with college completion. Both of these features are supported by the empirical literature (Card and Krueger 1992, Altonji 1993).

Labor Market Parameters. The expected unemployment duration in submarket s is given by $1/(\lambda_s^U + \mu)$.[36] Given μ, the unemployment duration of each education group in the data thus identifies the contact rate for the unemployed λ_s^U. However, λ_s^U is endogenous and pinned down by equation (11). The denominator on the right-hand side of that equation is calculated from the data on unemployment rates and educational attainment distribution. Thus, the unemployment duration effectively identifies the determinants of the firms' recruiting policies, which appear in the numerator of equation (11). In particular, the recruiting cost parameter χ in equation (17) is calibrated to match the average unemployment duration in the data. This gives $\chi = 0.0003$ for 1970 and 0.0108 for 2015. For the employed, we have $\lambda_s^E = (1 - \zeta)\lambda_s^U$ and set $\zeta = 1/3$,

[35] For example, Engbom and Moser (2017) use employer-employee matched data from Ohio between 2003 and 2012 and find that workers with bachelor's degrees earn 17.6% more in weekly wages than those with associate's degrees (their baseline group) after controlling for firm fixed effects and workers' characteristics. In our calibration, though not directly comparable to their estimate without controlling for worker fixed effects, the corresponding number (R_2/R_1) is 16.7% in 1970 and 17.5% in 2015.

[36] Unemployment can be interrupted either because the worker finds a job or because she dies. Since both events behave as Poisson processes, interruption is also Poisson with arrival rate $\lambda_s^U + \mu$.

which implies a monthly job-to-job transition rate of 1.9% in 1970 and 1.5% in 2015.[37] Given μ and λ_s^U, we use the unemployment rate of each education group and equation (13) to identify the separation hazard rate δ. This gives $(\delta_0, \delta_1, \delta_2) = (006, .010, .002)$ for 1970 and $(.013, .007, .001)$ for 2015, expressed as monthly frequencies.[38]

Firm Productivity and Bargaining. The firm productivity distribution is Pareto, $\Gamma(p) = 1 - p^{-\gamma}$, with Pareto parameter γ. We allow for different values of γ in 1970 and 2015 to capture time variation in firm productivity dispersion. The bargaining parameter β is time invariant and common across submarkets.

Intergenerational Linkage and Skill Distribution. We assume that ability a follows a symmetric Markov chain across generations and calibrate $\Pr[a = a_i \mid a_{-1} = a_i] = .66$ to match the correlation between a and a_{-1} in the data, which is obtained from Abbott et al. (2019) and is approximately 0.33.[39] We also assume that $\log z_0 \sim N(\mu_a, \sigma_a^2)$. The elasticity parameter α_a is time invariant. We thus assume that nature does not change over time but that nurture (endogenously) does.

This specification has two advantages. First, when the distribution of wages is Pareto log-normal, as is nearly the case in our calibration, the skill distribution also becomes Pareto log-normal, as typically assumed in the literature.[40] Second, we assume no ad hoc change in the innate skill distribution over time. Skill heterogeneity can vary only due to an endogenous change in the wage distribution.

Method of Simulated Moments. The model has two-sided heterogeneity of workers and firms. An ideal data set to estimate it would be an employer-employee matched data set of representative samples of workers and firms both in 1970 and 2015 that includes information on educational attainment of workers. To the best of our knowledge, it does not exist for the U.S.[41] We thus target between- and within-group inequality

[37]In Shimer (2005), the empirical estimates of the monthly job-to-job transition rate range from 2.5% to 3% based on monthly CPS data and from 0.5% to 2% based on March CPS data.

[38]We find the separation hazard rates that are not monotone in education levels in 1970, consistent with the non-monotone unemployment duration in 1970 in the data.

[39]Abbott et al. (2019) calculate transition probabilities based on test scores of cognitive abilities for mothers and children in the NLSY79. We use these probabilities to simulate a two-state Markov process and calculate the correlation in the simulated series.

[40]Denote the Pareto log-normal skill distribution as $z \sim PLN(\mu_z, \sigma_z^2, \lambda_z)$. This nests the standard distributions used in the literature, approaching log-normal when $\lambda_z \to \infty$ and Pareto when $\sigma_z \to 0$.

[41]As far as we know, at least two employer-employee matched data sets exist for the U.S., both of which have highly restricted access. One is the Master Earnings File (MEF) of the Social Security Administration. The MEF contains annual labor earnings information from 1978 for almost all workers in the U.S., but it does not contain information about their educational attainment. The other is the Longitudinal Employer-Household Dynamics (LEHD) of the U.S. Census Bureau. The LEHD program

Table 1: Model Performance

Targeted Moments	Data		Model	
	1970	2015	1970	2015
Between-group: SC-HS	15.2	22.4	8.8	19.9
Between-group: CL-HS	39.0	74.2	38.5	73.3
Within-group: HS	0.14	0.29	0.14	0.27
Within-group: SC	0.18	0.30	0.18	0.32
Within-group: CL	0.25	0.39	0.25	0.38

Note: Between-group inequality is measured as the difference in average log earnings for two educational groups. Within-group inequality is defined as the conditional variance of log earnings. Between-group inequality of SC relative to HS is not targeted.

measures and then argue why these wage moments are informative to identify key structural parameters, especially the parameter of the firm productivity distribution γ. We abstain from using firm-level data to directly estimate γ, because true firm productivity is never observed. We instead validate our choice of γ with firm-level data from Compustat by comparing directly estimable objects in the data to their model counterparts.

We set μ_{a_0} such that $\mathbb{E}z_0 = 1$ and calibrate the remaining 8 parameters $\vartheta \equiv \{\gamma_{70}, \gamma_{15}, \beta, \mu_{a_1}, \sigma_{a_0}, \sigma_{a_1}, \alpha_{a_0}, \alpha_{a_1}\}$. Since the two steady states share time-invariant parameters, we jointly calibrate the economy to 1970 and 2015 by solving

$$\min_{\vartheta \in \Theta} \left[\mathcal{M}(\vartheta) - \mathcal{M}_{\text{data}}\right]^{\mathrm{T}} \mathbf{I}_{||\mathcal{M}||} \left[\mathcal{M}(\vartheta) - \mathcal{M}_{\text{data}}\right], \tag{26}$$

where $\mathcal{M}_{\text{data}}$ is a vector of the 8 targeted empirical moments (see Table 1) and $\mathcal{M}(\vartheta)$ is a vector of the corresponding model moments given ϑ. We do not target the between-group inequality of SC relative to HS. We use the identity weighting matrix. Since the model is highly nonlinear, we apply a global optimization algorithm using quasi-random numbers generated by the Sobol sequence (see Appendix E).

Some comments are in order. The Pareto parameter γ_t of the firm productivity distribution drops from 31.7 to 9.9, implying a higher dispersion of firm productivity in 2015, which is externally validated in the next section. The bargaining parameter β is given by 0.27. We assume a constant β over time because it is difficult to identify the rent share directly in the data. However, this does not mean that the aggregate

contains detailed information about workers including their educational attainment. However, states participate in the program at different times and the majority of them participated only in late 90s.

labor share in the model is also constant over time; it indeed decreases by 12.7%, which is, though levels are not directly comparable, similar to the drop of 12.1% in the data.[42] Cahuc et al. (2006) estimate the bargaining parameter for different occupations and industries using French data. Excluding the exceptionally high estimate of 0.98 for executives, managers, and engineers in the construction sector, the remaining 15 estimates range from 0.00 to 0.38. In their preferred specification, Flinn and Mullins (2015) estimate the bargaining parameter of 0.252, which is close to our estimate, although the setups are not fully comparable. They also assume a common parameter across schooling levels as we do.

Finally, we obtain $(\mu_{a_0}, \mu_{a_1}, \sigma_{a_0}, \sigma_{a_1}) = (-.072, -.065, .36, .38)$, which gives $\mathbb{E}[z_0|a_0] < \mathbb{E}[z_0|a_1]$. These estimates, together with $(\alpha_{a_0}, \alpha_{a_1}) = (.37, .39)$, imply a positive correlation between z and a and, thus, positive education sorting.

5.3 Model Performance

Exactly Matched Moments. By choosing the corresponding parameters, the model replicates the educational attainment distributions, the unemployment rates, and the average unemployment durations in 1970 and 2015 (see Table A1).

Targeted Moments and Identification. Table 1 shows that the model well replicates the increasing trends in between- and within-group inequality measures defined in Section 2. Note that we can match the inequality measures not only for HS and CL but also for SC, a group that is often disregarded in the literature. Our model thus successfully captures key aspects of education choices and earnings distributions for different groups of the population.

The structural parameters ϑ are closely linked with the inequality measures, but formal identification is difficult. We instead develop an identification argument by means of numerical comparative statics. It is important to note that the time-invariant parameters (e.g., β) are identified, together with the time-variant parameter γ_t, by targeting the moments in both years.

First, α_{a_0} is a scaling parameter controlling the relative importance of innate skill and parental wages for ex ante labor productivity. A crucial parameter in our model is α_{a_1} or, more precisely, $\alpha_{a_1} - \alpha_{a_0}$. Increasing α_{a_1} while holding α_{a_0} fixed has two effects: (i) it increases the correlation between skill and ability and thus enhances the

[42]The level of labor share is generally lower in the data than in the model that abstracts from capital in the production. The data are annual labor share of nonfarm business sectors. Source: U.S. Bureau of Labor Statistics.

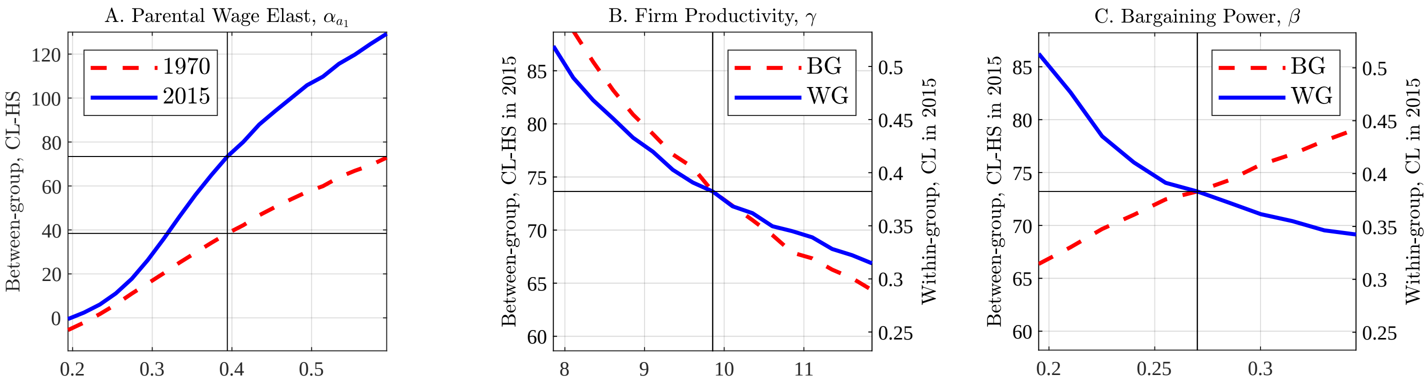

Figure 3: Identification. Panel A plots between-group inequality for CL in 1970 and 2015 against the parental wage elasticity α_{a_1}. Panels B and C plot between- and within-group inequality measures for CL in 2015 against the Pareto parameter γ and bargaining parameter β, respectively. The vertical and horizontal lines correspond to the calibrated parameter value and the value of the model moment, respectively.

degree of education sorting because workers' prior belief is more precise, and (ii) it increases the skill level of high-ability types and hence results in the high average skills of CL workers through education sorting. These effects lead to a larger worker skill difference between CL workers and HS workers.[43] This relationship is displayed in Panel A of Figure 3 in which we plot equilibrium between-group inequality against α_{a_1} for CL workers in both years, holding the other parameters in ϑ, including α_{a_0}, fixed.[44] This shows that the college premium is increasing in α_{a_1}, suggesting that the degree of education sorting cannot be too high or too low. Note that if α_{a_1} is sufficiently small, the college premium can even be negative due to negative sorting.

In Panels B and C of Figure 3, we plot both between- and within-group inequality for CL workers against the Pareto parameter γ and bargaining parameter β in 2015.[45] These plots show that within-group inequality decreases in both γ and β, while between-group inequality decreases in γ but increases in β. These moments jointly identify (γ, β); all else being equal, there exists no other pair (γ, β) that delivers the same values for between- and within-group inequality.

A higher Pareto parameter γ makes the right tail of the firm productivity distribu-

[43] As noted earlier in Section 4.3, stronger education sorting alone has an ambiguous effect on between-group inequality because it reduces the average skills of both HS and CL workers, which obscures the effect on the worker skill difference.

[44] We also obtain similar patterns when we hold α_{a_1} constant and decrease α_{a_0} instead (between-group inequality is decreasing in this case). See Appendix F.

[45] In all cases, we obtain very similar figures for 1970 and for SC workers (see Appendix F).

tion thinner, and thus, within-group inequality decreases mainly due to a smaller firm effect. Between-group inequality also decreases due to a smaller firm pay difference. This is because the thinner right tail of the p distribution weakens firm sorting; the vacancies created by very productive firms decrease, and this effect is more pronounced in the CL market.

A higher bargaining parameter β makes wages less dependent on outside offers (equation 14), and thus, within-group inequality decreases due to a smaller friction effect. In addition, when β is high, earning a college degree becomes relatively more attractive because the gap between the unemployment values of HS and CL increases (equation 18). Workers receive a larger share of the surplus, which pushes up wages (equation 14) and, because of intergenerational linkages, shifts the distribution of pre-college skills. Since these workers are more likely to finish college due to stronger education sorting, between-group inequality increases.

External Validation. Our calibration is externally validated by various empirical moments that are not targeted. The success in matching these moments lends credence to the model's predictions on the wage inequality trend.

First, since the main driver of increasing earnings inequality is an increase in firm productivity dispersion, it is paramount to validate our choice of γ using firm-level data. Using Compustat, in Section 2, we showed that the standard deviation of log output per worker increases by a factor of 2.1 from 1970 to 2015.[46] Reassuringly, this measure in the model increases similarly by a factor of 2.6[47]. In Figure 4, we also plot the distributions of firm size and output relative to their means in 2015 (see Appendix F for the figure for 1970). The model does a good job in replicating the overall shape of these distributions, especially, in capturing the significant fraction of small firms in terms of their size and output.

For the earnings distribution, Kelley's skewness increases from 0.26 to 0.43 in the data. In the model, that increases similarly from 0.32 to 0.41. Our model also qualitatively matches the trends in between- versus within-firm inequality. Song et al. (2019) find that two-thirds of the rise in variance of log earnings between 1981 to 2013 occurred due to a rise in the inequality between firms. Similarly, in our model, it explains the majority (59%) of the overall increase.

The model displays positive education sorting. How can we assess whether its degree

[46]The standard deviation of log value added per worker increases by a factor of 2.6.

[47]The model counterpart of observed output per worker depends on firm productivity and on the distribution of employment across firms, both within and across submarkets.

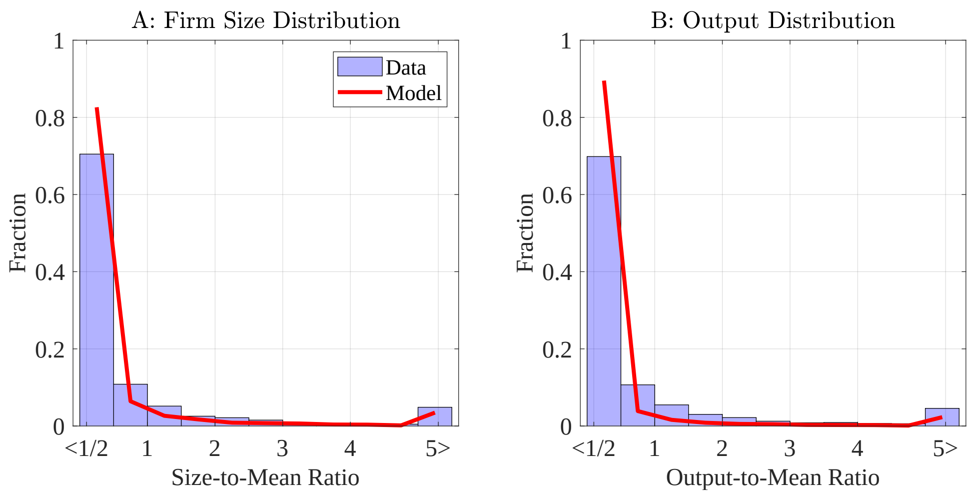

Figure 4: Firm Size and Output Distribution. The figure plots the distributions of firm size (Panel A) and output (Panel B) in the data and the model in 2015. The values are normalized relative to mean; e.g., the leftmost bar in Panel A shows that the fraction of firms whose size is below half of the mean size.

is reasonable? Using NLSY data from 1979-90, Belzil and Hansen (2002) estimate the correlation between unobserved skill and years of schooling of 0.28. We compute the correlation between skill and years of schooling using the college duration parameter in our model, and we obtain 0.31 for 1970, which is close to their estimate.[48]

The model also displays positive firm sorting; high-productivity firms tend to be large and recruit CL workers more intensively, whereas low-productivity firms tend to be small and recruit HS workers more intensively. Using the Quarterly Workforce Indicators, we find two pieces of indirect evidence of firm sorting.[49,50] First, small firms hire more HS workers than CL workers, but large firms do the opposite. In particular, the shares of HS and CL workers working at firms with fewer than 50 workers are 36% and 29%, respectively, whereas those working at firms with at least 500 workers are 31% and 35%, respectively.[51] Second, HS workers are more likely to work at small

[48]Interestingly, our model indicates that the correlation increases to 0.37 in 2015.

[49]The Quarterly Workforce Indicators are aggregated data of LEHD. The data are available from 1993 to 2017. We use male workers and focus on HS, SC, and CL only for the analysis.

[50]Establishing more direct evidence of firm sorting is challenging for two reasons. First, firm productivity is not directly observed. Second, we have to have data for a representative sample of the U.S. firms on their vacancy creation and educational requirements for each vacancy, which do not seem to exist.

[51]In line with this argument, the shares of HS and CL workers working at firms with 50-249 workers are 35% and 29%, respectively, and those working at firms with 250-499 workers are 34% and 31%, respectively.

firms than CL workers, and CL workers are more likely to work at large firms than HS workers. Specifically, among HS workers, 33% work at firms with fewer than 50 workers, and 44% work at firms with at least 500 workers. The corresponding numbers for CL workers are 28% and 52%.

The degree of sorting increases over time in the model: the rank correlation between skill and schooling attainment increases from 0.29 to 0.36, and that between the productivity of vacancies and the submarkets in which these vacancies are posted also increases from 0.08 to 0.20.[52] These results imply an increase in the extent of assortative matching. How does it contribute to overall inequality? In our model, assortative matching explains 4.0% of the variance of log earnings in 1970 and 11.8% in 2015.[53] This increase is comparable to that in Song et al. (2019), who report 4.7% for the period 1980-86 and 11.7% for the period 2007-13. Therefore, like that paper, we find that an increase in assortative matching is one of the main causes of increasing inequality.

Our model also matches the estimated elasticities of college attendance and graduation rates with respect to financial aid in Castleman and Long (2016). They examine the effects of need-based grants in Florida using a regression-discontinuity design and find that a $\$1,300$ increase in yearly grants for children with parental income of $\$30,000$ increases enrollment by 12% and graduation within 5 years by 20%. Simulating a similar experiment in the model for 2015, we find elasticities of 17% and 31%, respectively, which are somewhat larger but in line with Castleman and Long (2016).[54]

Finally, empirically validating the concept of the intergenerational linkages we consider in equation (1) is paramount because it is admittedly highly stylized. Also, the intergenerational linkages via initial skill formation are a mechanism key to explaining the increasing inequality observed in our model (see Section 6). To empirically validate it, we check the intergenerational mobility implied by the model. Specifically, Chetty et al. (2014) estimate the rank-rank slope, i.e., the correlation between child and parent percentile ranks, as a measure of intergenerational mobility. Figure 5 reproduces their Figure IIa, a binned scatter plot of the mean percentile income rank of children against their parents' percentile income rank, and adds a corresponding model-generated plot

[52] The Spearman rank correlation is defined by $corr(r_i, r_j)$, where r_i and r_j are the ranking of variables i and j.

[53] The standard definition is the joint variability of worker and firm fixed effects and is given by $2\text{Cov}[\mathbb{E}(\ln z \mid p), \mathbb{E}(\ln \phi \mid p)]$. See Song et al. (2019).

[54] In the model, we do this experiment by reducing the annual costs of college by $\$1,300$ for families with $\$30,000$ of income.

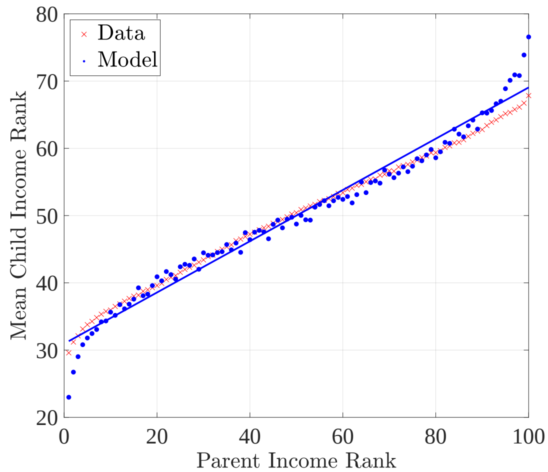

Figure 5: Intergenerational Mobility. The figure presents a binned scatter plot of the mean percentile income rank of children against their parents' percentile income rank, with a line fitted to the model. The data are taken from Chetty et al. (2014).

with a fitted line obtained by simulating the model in 2015, which is comparable to their baseline time frame.[55] Reassuringly, these plots are quite similar. Our simulated rank-rank slope is 0.381, which is close to their estimate of 0.341.

6 Quantitative Analysis

The results of the previous section indicate that to match the data, the model requires the firm productivity dispersion in 2015 to be higher than that in 1970—a smaller Pareto parameter γ of the firm productivity distribution in 2015. In this section, we analyze how this increase in firm productivity dispersion affects between- and within-group inequality from 1970 to 2015. For exposition purposes, we focus on HS and CL workers. The results for SC workers are available in Appendix F.

To understand the channels through which increasing dispersion affects inequality, we conduct two counterfactual experiments in which we take the 1970 economy and increase the dispersion of firm productivity by setting γ to the (low) level of 2015. In the first experiment, we shut down skill supply responses by holding fixed the precollege skill distribution and the schooling attainment distribution to those in 1970. In

[55]The core sample in Chetty et al. (2014) consists of children in the 1980-82 birth cohort and their parents. Income is measured as mean income in a relatively short period of time (1996-2000 for parents and 2011-2012 for children). The paper shows that the baseline measures do not suffer from significant lifecycle or attenuation bias. Similarly, in our model, we construct a stationary (point-in-time) income distribution for parent workers and that for their children.

Table 2: Counterfactual Analysis and Decomposition: Between-Group Inequality

	[1] 1970	[2] Ψ_{70}	[3] Z_{70}	[4] 2015
Between-group: CL-HS	**38.4**	**61.8**	**52.3**	**65.2**
Returns to Education	18.2	18.2	18.2	18.9
Worker Skill Difference	11.4	11.4	9.1	21.1
Firm Pay Difference	8.9	32.2	25.1	25.2

Note: The inequality measures in 1970 and 2015 are different from those in Table 1 as the time-variant labor-market parameters (χ and δ) are set at the average levels between 1970 and 2015.

the second, we still hold fixed the precollege skill distribution, but allow for endogenous skill supply responses through college education decisions. To isolate the effect of a change in skill demand via increasing productivity dispersion and skill supply responses, throughout the experiments, we hold fixed the other time-variant labor-market parameters, i.e., χ and δ, at their average levels between 1970 and 2015. Thus, the inequality measures in 1970 and 2015 in Tables 2 and 3 differ slightly from those reported in Table 1.

These experiments are motivated as follows. In our model, an exogenous increase in productivity dispersion affects inequality not only directly through marginal products but also indirectly by triggering a change in the demand for skills because firms adjust their vacancy posting behavior. In the first experiment, we isolate these two effects by shutting down skill supply responses, which highlights the role of skill supply and schooling decisions in shaping inequality. In the second experiment, we activate the response of the skill supply through optimal education decisions, which highlights the role of intergenerational linkages. By comparing these experiments with the economies in 1970 and 2015, we can understand the effects of the increasing dispersion in firm productivity on inequality.

6.1 Between-Group Inequality

Table 2 reports the results of between-group inequality for CL relative to HS (i.e., the college premium). The columns labeled "[2] Ψ_{70}" and "[3] Z_{70}" present the results for the first and second counterfactual experiments, respectively.

Absent skill supply responses, between-group inequality increases from 38.4 log points in column [1] to 61.8 log points in column [2], accounting for 87% of the observed increase between 1970 and 2015. Column [3] shows that allowing for optimal college education choices attenuates this increase; between-group inequality drops to 52.3 log points, which accounts for only 52% of the actual increase. Finally, when we also allow for intergenerational linkages, between-group inequality increases to 65.2 log points in 2015.

Overall, this indicates that the response of the skill supply has sizable effects on between-group inequality. The effect of the increase in productivity dispersion on this measure is attenuated by endogenous education decisions, as the change from column [2] to [3] demonstrates. In contrast, it is amplified by intergenerational linkages, as the change from column [3] to [4] demonstrates. The comparison between columns [2] and [4] suggests that these two forces almost cancel one another out in equilibrium.[56]

Decomposition of Between-Group Inequality. To elucidate the economic forces behind these results, we decompose between-group inequality into the three elements shown in equation (24): returns to education, the worker skill difference, and the firm pay difference.

The comparison between 1970 and 2015 indicates that pure returns to education increase only slightly. Their contribution to the college premium falls from 47% to 29%, shifting from being the largest in 1970 to become the smallest in 2015. The worker skill difference increases from 11.4 log points to 21.1 log points, which implies that the difference in average skills between CL and HS workers is almost twice as large in 2015 than in 1970. The firm pay difference also increases from 8.9 log points to 25.2 log points, becoming the largest component of the college premium in 2015. These changes in the worker skill difference and the firm pay difference account for 36% and 61%, respectively, of the increase in between-group inequality (+26.8 log points); returns to education play a very limited role in explaining this change.

From column [1] to [2], the increase in between-group inequality is, by construction, due exclusively to a larger firm pay difference, which is now 32.2 log points. With the value of γ in 2015, there are more high-productivity firms that have high demand for skills. Wages increase in all submarkets but disproportionately more in the CL market, because these high-productivity firms intensively recruit CL workers, resulting in stronger firm sorting and a larger firm pay difference.

This force is mitigated once we allow for optimal college education decisions (column [3]), because both worker skill difference and firm pay difference decrease. A higher college premium makes earning a college degree more attractive, which increases the

[56] We find a qualitatively similar pattern for between-group inequality of SC relative to HS (Appendix F).

supply of CL workers.[57] The workers for which educational attainment changes between columns [2] and [3] tend to be less skilled than the average CL worker but more skilled than the average HS worker, and thus the average skill of each group is lower in column [3]. The response of the worker skill difference depends on the change in the average skill of CL workers relative to that of HS workers. In this case, the former decreases more than the latter, resulting in a smaller worker skill difference.

Also, column [3] shows that the firm pay difference decreases by 7.1 log points. Since a larger (smaller) supply of CL (HS) workers makes it easier (harder) for firms to contact these workers, the logic of competitive labor markets suggests that the firm pay difference should fall in response. However, it is important to note that there is also a countervailing effect in the presence of labor market frictions. From column [2] to [3], the average skill of CL workers decreases, and thus high-productivity firms cut down their vacancy creation (see equations 16-17). Workers now expect a smaller option value of climbing-the-ladder and demand even higher wages at low-productivity firms. This distributional effect of high-productivity firms' recruiting policies on the option value of all low-productivity firms thus countervails partly, but not fully, the negative impact of a larger supply of CL workers on the firm pay difference.

Comparing columns [3] and [4], we observe an increase in between-group inequality, which is mainly due to the increase in the worker skill difference from 9.1 log points in column [3] to 21.1 log points in column [4]. This change occurs mainly for two reasons: the change in the precollege skill distribution and the stronger education sorting. In column [4], the presence of intergenerational linkages translates the higher dispersion of the wage distribution into a higher dispersion of the precollege skill distribution. Moreover, with the value of γ in 2015, the right tail of the wage distribution becomes thicker. Since the precollege skill distribution inherits this property, this results in stronger education sorting because workers in the upper end of this distribution are more likely to finish college.

6.2 Within-Group Inequality

Table 3 reports the results of within-group inequality for HS and CL workers. Absent skill supply responses, within-group inequality increases from 0.24 in column [1] to 0.38 in column [2] for CL workers, constituting 93% of the observed increase from 1970 to 2015, and from 0.13 to 0.25 for HS workers, constituting 69% of the observed increase.

[57] The fraction of CL workers is 28.4% in columns [1-2], 36.8% in column [3], and 39.8% in column [4]. The corresponding numbers for HS workers are 54.6%, 33.1%, and 31.0%, respectively.

Table 3: **Counterfactual Analysis and Decomposition: Within-Group Inequality**

	[1] **1970**	[2] Ψ_{70}	[3] Z_{70}	[4] **2015**
A. Within-group HS	**0.13**	**0.25**	**0.27**	**0.30**
Person effect	0.13	0.13	0.13	0.15
Firm effect	0.00	0.07	0.07	0.08
Friction effect	0.00	0.06	0.07	0.07
B. Within-group CL	**0.24**	**0.38**	**0.36**	**0.40**
Person effect	0.23	0.23	0.21	0.24
Firm effect	0.01	0.09	0.09	0.10
Friction effect	0.00	0.06	0.06	0.06

Note: The inequality measures in 1970 and 2015 are different from those in Table 1 as the time-variant labor-market parameters (χ and δ) are set at the average levels between 1970 and 2015.

In contrast to between-group inequality, however, allowing for college education choices in column [3] barely mitigates this change; within-group inequality remains at 0.36 for CL workers and to 0.27 for HS workers. The change from column [3] to [4] is also somewhat modest. For example, for CL workers, allowing for intergenerational linkages increases within-group inequality by 0.04, which is 22% of the overall increase.[58]

These results suggest that the change in within-group inequality is mainly driven by the shock to the firm productivity dispersion and associated responses of skill demand. In contrast to between-group inequality, the responses of skill supply, especially those via endogenous college decisions, have modest effects and thus play a limited role in determining within-group inequality.

Decomposition of Within-Group Inequality. To understand the economic forces behind these results, we decompose within-group inequality into the three elements shown in equation (25): the person effect, the firm effect and the friction effect.

The comparison between 1970 and 2015 shows that the increase in within-group inequality over time is mainly due to an increase in the firm and the friction effect. The firm effect explains only 3.2% of within-group variation for HS and 2.8% for CL in 1970, but 26.0% for HS and 24.0% for CL in 2015. The friction effect explains 2.7% of within-group variation for HS and 1.6% for CL in 1970, but 23.9% for HS and 15.9% for CL in 2015. Although the person effect explains the largest part of within-group inequality for both education groups at any point in time (94.0% for HS and 95.6% for

[58]We find a qualitatively similar pattern for within-group inequality of SC workers (Appendix F).

CL in 1970; 50.1% for HS and 60.1% for CL in 2015), it remains relatively constant over time, and thus it does not contribute to the increase in within-group inequality.[59]

The change in both the firm and the friction effect are closely linked to the Pareto parameter γ. A lower γ means that the firm productivity distribution is more dispersed, which directly translates into a higher firm effect. This is because wages per efficiency unit of labor ϕ^s depend on the productivity of the employer (equation 14), and thus, wage variation across firms hinges on the shape of the productivity distribution. In addition, the appearance of many high-productivity firms increase the variation of outside offers, and thus climbing the ladder within less productive firms becomes more likely, which results in higher frictional wage variation. Hence, the value of the firm and friction effects are high in columns [2-4], where the firm productivity dispersion is larger.

Comparing columns [2] and [3] with [4] shows that the increase in firm productivity dispersion is attenuated via endogenous college decisions and amplified via intergenerational linkages, which appears in the change in the person effect especially for CL. However, the effect is small in contrast to the case of between-group inequality.[60]

7 Conclusions

In this paper, we develop an equilibrium model with endogenous skill demand via firms' recruiting policies and skill supply via college education investment and intergenerational linkages. The model features education sorting of workers and labor market sorting of firms, and it provides novel decomposition equations that can be used to study how the increase in firm productivity dispersion affects between- and within-

[59]Our decomposition results align well with previous literature. First, the result suggests that wage variation within education groups is largely due to skill variation among workers, especially for CL workers, which is similar to Lee et al. (2017). Second, the person effect is more important for explaining wage variation among highly educated workers. This point is related to Postel-Vinay and Robin (2002), who find that the person effect is more important for high-skilled occupations. Finally, consistent with Chen (2008), we find that the large gap in within-group inequality between HS and CL is mostly explained by the difference in worker heterogeneity, suggesting that the skill heterogeneity is more dispersed among CL workers.

[60]In the second experiment (column [3]), earning a college degree is more attractive because of the higher college premium; thus, the marginal students who tend to be less skilled than the average college worker are more likely to become college graduates. The optimal prior cutoff $\bar{q}$ indeed decreases compared to the economy in 1970 or column [2]. However, this does not necessarily translate into a higher skill variance for CL. This is because the skill variation (i.e., person effect) after the marginal students change their educational attainment depends on their relative density. If their density is relatively large, then this change could result in a reduction of the person effect, which indeed occurs in our experiment.

group inequality over time in the U.S.

We now highlight two important lessons from our analysis that should be useful for economic research and practical policies. First, we show that the effect of the increase in firm productivity dispersion is attenuated by endogenous education decisions but amplified by intergenerational linkages. This means that the interaction of demand and supply of skills plays a crucial role in shaping wage inequality, especially for between-group inequality, and thus, these equilibrium forces must be explicitly considered in any study of the sources of inequality. Second, any policy to address the problem of growing inequality must be carefully designed to take into account this complex interaction of demand and supply of skills. For example, a policy promoting educational attainment would affect the sorting of workers and, hence, the sorting of firms in equilibrium and might thus have unintended consequences.

Our model can be enriched along several dimensions. First, we only consider human capital investments prior to labor market entry. While education is arguably one of the most important forms of investment, workers can also invest in their productivity while working through, e.g., on-the-job training or learning-by-doing. Such productivity enhancements may have interesting interactions with labor market frictions. Second, for sake of tractability, in our model, a single parameter represents all kinds of costs associated with education, including financial costs and psychic costs. However, it would be important to explicitly model these costs if one wants to conduct policy experiments in the model. We leave these extensions for future work.

References

Abbott, B., G. Gallipoli, C. Meghir, and G. Violante (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy 127*(6), 2569–2624.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High Wage Workers and High Wage Firms. *Econometrica 67*(2), 251–333.

Acemoglu, D. (1998). Why do new technologies complement skills? directed technical change and wage inequality. *The Quarterly Journal of Economics 113*(4), 1055–1089.

Acemoglu, D. (1999). Changes in unemployment and wage inequality: An alternative theory and some evidence. *American Economic Review 89*(5), 1259–1278.

Altonji, J. G. (1993). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics 11*(1), 48–83.

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.

Bagger, J. and R. Lentz (2018). An empirical model of wage dispersion with sorting. *The Review of Economic Studies 86*(1), 153–190.

Barth, E., A. Bryson, J. C. Davis, and R. Freeman (2016). It's where you work: Increases in the dispersion of earnings across establishments and individuals in the united states. *Journal of Labor Economics 34*, S67–S97.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press.

Belzil, C. and J. Hansen (2002). Unobserved Ability and the Return to Schooling. *Econometrica 70*(5), 2075–2091.

Brynjolfsson, E. and A. Saunders (2010). *Wired for Innovation: How Information Technology Is Reshaping the Economy*. MIT Press, Cambridge.

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. Volume 3 of *Handbook of Labor Economics*, pp. 1801–1863. Elsevier.

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.

Castleman, B. L. and B. T. Long (2016). Looking beyond enrollment: The causal effect of need-based grants on college access, persistence, and graduation. *Journal of Labor Economics 34*(4), 1023–1073.

Chen, S. H. (2008). Estimating the variance of wages in the presence of selection and unobserved heterogeneity. *The Review of Economics and Statistics 90*(2), 275–289.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics 129*(4), 1553–1623.

Decker, R. A., J. C. Haltiwanger, R. S. Jarmin, and J. Miranda (2018). Changing business dynamism and productivity: Shocks vs. responsiveness. Working Paper 24236, National Bureau of Economic Research.

Dunne, T., L. Foster, J. Haltiwanger, and K. R. Troske (2004). Wage and productivity dispersion in united states manufacturing: The role of computer investment. *Journal of Labor Economics 22*(2), 397–429.

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.

Faggio, G., K. G. Salvanes, and J. Van Reenen (2010). The evolution of inequality in productivity and wages: panel data evidence. *Industrial and Corporate Change 19*(6), 1919–1951.

Flinn, C. and J. Mullins (2015). Labor market search and schooling investment. *International Economic Review 56*(2), 359–398.

Flood, S., M. King, R. Rodgers, S. Ruggles, and J. R. Warre. Integrated Public Use Microdata Series, Current Population Survey: Version 6.0 [dataset]. Minneapolis, MN: IPUMS, 2018. https://doi.org/10.18128/D030.V6.0.

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.

Hagedorn, M., T. H. Law, and I. Manovskii (2017). Identifying equilibrium models of labor market sorting. *Econometrica 85*(1), 29–65.

Hartman-Glaser, B., H. N. Lustig, and M. Z. Xiaolan (2019). Capital share dynamics when firms insure workers. *Journal of Finance 74*(4), 1707–1751.

Heckman, J., L. Lochner, and C. Taber (1998). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics 1*(1), 1–58.

Heckman, J. J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science 312*(5782), 1900–1902.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). *Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond*, Volume 1 of *Handbook of the Economics of Education*, Chapter 7, pp. 307–458. Elsevier.

Helpman, E., O. Itskhoki, and S. Redding (2010). Inequality and unemployment in a global economy. *Econometrica 78*(4), 1239–1283.

Hendricks, L., C. Herrington, and T. Schoellman (forthcoming). College quality and attendance patterns: A long-run view. *American Economic Journal: Macroeconomics*.

Hendricks, L. and O. Leukhina (2017). How risky is college investment? *Review of Economic Dynamics 26*, 140–163.

Hendricks, L. and T. Schoellman (2014). Student abilities during the expansion of US education. *Journal of Monetary Economics 63*(2), 19–36.

Huggett, M., G. Ventura, and A. Yaron (2011). Sources of lifetime inequality. *American Economic Review 101*(7), 2923–54.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Keane, M. P. and K. I. Wolpin (1997). The career decisions of young men. *Journal of Political Economy 105*(3), 473–522.

Kehrig, M. (2015). The cyclical nature of the productivity distribution. Working Paper 25114.

Krueger, D. and A. Ludwig (2016). On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium. *Journal of Monetary Economics 77*, 72–98.

Krusell, P., L. E. Ohanian, J.-V. Rìos-Rull, and G. L. Violante (2000). Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica 68*(5), 1029–1053.

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.

Lise, J., C. Meghir, and J.-M. Robin (2016). Matching, sorting and wages. *Review of Economic Dynamics 19*, 63–87.

Lopes de Melo, R. (2018). Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy 126*(1), 313–346.

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.

Papageorge, N. W. and K. Thom (forthcoming). Genes, education, and labor market outcomes: Evidence from the health and retirement study. *Journal of the European Economic Association*.

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

Shimer, R. (2005). The cyclicality of hires, separations, and job-to-job transitions. *Federal Reserve Bank of St. Louis Review 87*(4), 493–508.

Song, J., D. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019). Firming up inequality. *The Quarterly Journal of Economics 134*(1), 1–50.

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.

Taber, C. (2001). The rising college premium in the eighties: Return to college or return to unobserved ability? *The Review of Economic Studies 68*(3), 665–691.

Inequality, Productivity Dispersion, and Supply of Skills

Online Appendix

Manuel Macera and Hitoshi Tsujiyama

A Data

A.1 Current Population Survey (CPS)

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

We first define the education categories. To do so, we use a variable for educational attainment provided by IPUMS-CPS.[1] We define high school dropouts as those with fewer than 12 years of completed schooling or those without a 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; and college graduates as those with 4 or more years of completed schooling. We do not use high school dropouts in the analyses. We measure experience as years of schooling subtracted from age minus 5. In the Mincerian regressions, we include cubic controls for experience.

For earnings 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 workers are excluded. All amounts are adjusted to 2015 U.S. dollars using the CPI. We impute average hourly wages for each observation using reported work weeks and

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

usual hours per week. We then drop those with imputed hourly wages falling below one-half of the federal minimum wage.

We follow Autor et al. (2008) for top coding. Prior to 1988, wage and salary incomes were collected in a single variable. After 1988, they were reported as 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 in 1996 or later. For those, top-coded values are assigned the mean of all top-coded earners, and 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.

A.2 Compustat

We use data of U.S. publicly traded firms from the Compustat North America Fundamentals Annual data set. The data set contains accounting data for all publicly traded firms since 1960. We focus on U.S. firms only and use data from 1968. We exclude financial firms (SIC codes from 6000-6799) and firms whose sales or number of employees are not positive. When constructing value added, we also exclude firms whose total assets or staff expenses are negative. All of the variables are winsorized at the 1% level.

We measure firm productivity as log output per worker (Decker et al. 2018), and we use sales data for calculating output. Sales data are available for about 94% of firm-year observations. The standard deviation of this measure relative to that in 1970 is reported in Figure 2. As a cross check, we also winsorize the productivity at the 1% or 5% level each year to avoid the influence of outliers, but the results are materially unchanged. In addition, we use revenue instead of sales (Decker et al. 2018), but the results are virtually identical.[2]

We also use log value added per worker to measure firm productivity, perhaps a more conventional definition of labor productivity (e.g., Hartman-Glaser et al. 2019). Value added is given by the sum of capital income and labor costs. Capital income is given as operating income before depreciation (OIBDP), which is equal to sales minus operating expenses including the cost of goods sold, labor costs, and other administrative expenses. Labor costs are given by staff expenses (XLR). However, since public firms are not required to file staff expenses, XLR is only reported for about

[2]These results are available upon request.

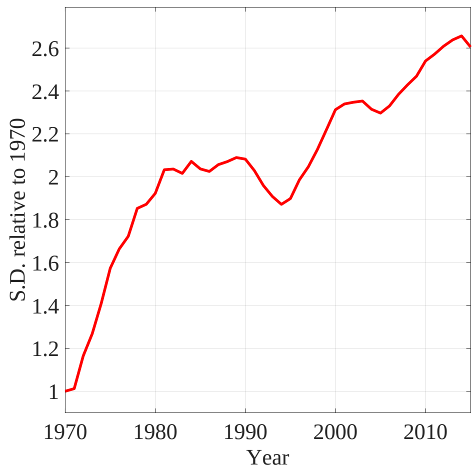

Figure A1: Productivity Dispersion Using Value Added. The figure plots the standard deviation of log value added per worker relative to that in 1970. All values are five-year-centered moving-averaged. Value added is adjusted to address negative values.

10.4% of firm-year observations. We thus closely follow the imputation procedure of Hartman-Glaser et al. (2019) and construct the extended labor costs for firms that fail to report XLR. First, we estimate the average labor cost per employee within the industry-size group for each year. We use the Fama-French 17-industry classification and then sort firms into 20 size groups within each industry based on their real total assets, yielding 340 industry-size cells.[3] The average labor cost per employee within each cell for each year is estimated, using the available XLR observations. Next, we use this estimate to impute labor costs for firms that have missing XLR as the number of employees multiplied by the average labor cost per employee of the same industry-size cell during that year. We winsorize the extended XLR at the 5% level each year to exclude outliers from the imputation. Value added is the sum of OIBDP and extended XLR, and we further winsorize value added at the 5% level each year to avoid the influence of outliers. Finally, we measure firm productivity as log value added per worker. The standard deviation of this measure relative to that in 1970 is reported in Figure A1. Note that, when taking log, we estimate the adjusted value added to address negative values as in Hartman-Glaser et al. (2019). Specifically, we identify the minimum operating income (OIBDP) for each year and then increase the value added of all firms by the absolute value of the minimum OIBDP times 101%. This adjustment does not affect the calculation of standard deviation.

[3]Since we exclude financial firms, we have effectively 320 cells.

B Further Discussion of the Model

Discrete Returns to Education. We assume that 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, also known as a "sheepskin effect", is empirically supported (e.g., Jaeger and Page 1996).

Wage Bargaining. Wage bargaining is crucial when analyzing sources of wage inequality. In our model, identical workers at the same firm may earn different wages. This wage dispersion stems from the different histories of offers these workers have received, and its source is wage increases within the firm ("climbing the ladder"), one of the central features of wage bargaining models pioneered by Postel-Vinay and Robin (2002).[4]

Risk Preferences. We assume that workers are risk neutral, which greatly simplifies the analysis. Since within-group inequality has increased over the past 30 years, one could argue that using risk-averse preferences would be more appropriate when studying the education choice. In the data, within-group inequality increased for all education levels, and the size of the increase is comparable; i.e., the variance of log wages has increased across the board by approximately 60% (see Figure 1). We thus argue that the education decision would be similar even if we used risk-averse preferences.

C Proofs

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

Proof of Proposition 1. The proof is by guess and verify. Rearranging equation (5) and applying integration by parts, we obtain

$$(\rho + \delta + \lambda^E)\pi(w, p) = \rho(p - w) - \lambda^E \int_{g(w,p)}^{p} \pi_w(\phi(x, p), p)\phi_1(x, p)F(x)dx. \tag{A1}$$

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

$$(\rho + \delta + \lambda^E)\pi_w(w, p) = -\rho + \lambda^E \phi_1(g(w, p), p)g_w(w, p)\pi_w(w, p)F(g(w, p)).$$

[4]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 an artifact of using a discrete firm productivity distribution, and with a continuous distribution, the dispersion vanishes.

Since, by the definition of $g(w,p)$, we have $\phi_1(g(w,p),p)g_w(w,p) = 1$, we can write

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

Plugging this expression back into equation (A1) and noting

$$\rho(p - w) = \rho \int_{g(w,p)}^{p} \phi_1(x,p)dx,$$

we have

$$\pi(w,p) = \rho \int_{g(w,p)}^{p} \frac{\phi_1(x,p)}{\rho + \delta + \lambda^E(1 - F(x))} dx. \tag{A2}$$

From equation (14), 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 (A2) yields equation (15). ■

Proof of Proposition 3. Using equations (8-9), it is immediate that both U and W are linear in z. This gives

$$\begin{aligned}
\rho W(w,p) &= \rho w + \delta \left[U - W(w,p)\right] + \lambda^E \int_{g(w,p)}^{p} \left[W(\phi(x,p),p) - W(w,p)\right] dF(x) \\
&\quad + \lambda^E \int_{p}^{\overline{p}} \left[W(\phi(p,x),x) - W(w,p)\right] dF(x), && \text{(A3)} \\
\rho U &= \rho b + \lambda^U \int_{p_{\inf}}^{\bar{p}} \left[W(\phi_0(x),x) - U\right] dF(x). && \text{(A4)}
\end{aligned}$$

The rent-sharing condition states that when an unemployed worker is matched with a p-firm, she is offered $\phi_0(p)$ and obtains

$$W(\phi_0(p),p) = U + \beta \left[W(p,p) - U\right].$$

Plugging this into equation (A4) and rearranging terms, we obtain

$$(\rho + \lambda^U \beta)U = \rho b + \lambda^U \beta \int_{p_{\inf}}^{\bar{p}} W(x,x)dF^s(x). \tag{A5}$$

Similarly, when a worker employed by a p-firm is poached by a p'-firm ($p' > p$),

$$W(\phi(p,p'),p') = W(p,p) + \beta\left[W(p',p') - W(p,p)\right].$$

Now, consider the case of $w = p$ in equation (A3). Noting that $g(p,p) = p$ and using the above expression, we obtain

$$\left[\rho + \delta + \lambda^E \beta(1 - F(p))\right] W(p,p) = \rho p + \delta U + \lambda^E \beta \int_p^{\overline{p}} W(x,x) dF(x).$$

Alternatively, using $\tilde{W}(p) \equiv W(p,p)$ and adding the terms in equation (A5),

$$\left[\rho + \delta + \lambda^E \beta(1 - F(p))\right] \tilde{W}(p) = \rho(p-b) + (\rho+\delta+\lambda^U\beta)U - \lambda^U \beta \int_{p_{\inf}}^{\overline{p}} \tilde{W}(x) dF(x) + \lambda^E \beta \int_p^{\overline{p}} \tilde{W}(x) dF(x).$$

Applying integration by parts, we obtain

$$\begin{aligned}(\rho + \delta + \lambda^E \beta)\tilde{W}(p) \;&=\; \rho(p-b) + (\rho + \delta + \lambda^U \beta) U \\ &\quad -\lambda^U \beta \left[\tilde{W}(\overline{p}) - \int_{p_{\inf}}^{\overline{p}} \tilde{W}'(x) F(x) dx\right] + \lambda^E \beta \left[\tilde{W}(\overline{p}) - \int_p^{\overline{p}} \tilde{W}'(x) F(x) dx\right].\end{aligned}$$

Taking the derivative of both sides with respect to p yields

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

Plugging this back in and rearranging terms,

$$\begin{aligned}(\rho + \delta + \lambda^E \beta)\tilde{W}(p) \;&=\; \rho(p-b) + (\rho + \delta + \lambda^U \beta) U + (\lambda^E - \lambda^U)\beta \tilde{W}(\overline{p}) \\ &\quad +\lambda^U \beta \int_{p_{\inf}}^{\overline{p}} \frac{\rho}{\rho + \delta + \lambda^E \beta(1 - F(x))} F(x) dx - \lambda^E \beta \int_p^{\overline{p}} \frac{\rho}{\rho + \delta + \lambda^E \beta(1 - F(x))} F(x) dx.\end{aligned}$$

Making $p = \overline{p}$, we obtain the expression for $\tilde{W}(\overline{p})$:

$$\tilde{W}(\overline{p}) = U + \frac{\rho}{\rho + \delta + \lambda^U \beta} \int_{p_{\inf}}^{\overline{p}} \frac{\rho + \delta + \lambda^E \beta(1 - F(x)) + \lambda^U \beta F(x)}{\rho + \delta + \lambda^E \beta(1 - F(x))} dx.$$

Again, plugging this back in and rearranging terms, we obtain

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

Using this equation in equation (A5), we obtain

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

Finally, applying integration by parts to the second term yields equation (18). ■

D Stationary Distribution

In this section, we characterize the stationary distribution of workers in closed form. We have numerically verified the following expressions using a discrete-time approximation of our continuous-time economy.

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

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

where $\psi_a(\cdot)$ is the probability density of skills conditional on $a \in \mathcal{A}$. Define $q(t)$ as the belief at time t without any completion shock; thus, $q(0) = q_0$. Solving the differential equation (3), we obtain

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

where $\Delta\theta^s \equiv \theta_1^s - \theta_0^s$. We define a function $t(q, q') = \{t' - t | q = q(t), q' = q(t')\}$ for the length of time for which the belief falls from q to q'. This can be solved as

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

Using this expression, we can solve for the mass of workers in each submarket that only depends on the optimal belief cutoffs, $\{\overline{q}^s\}_{s\in\{1,\cdots,S\}}$.

We focus on the case in which $S = 2$ for illustration purposes. Given the following expressions for the density of workers $\hat{\Psi}_a^s$, the (conditional) stationary distribution

$\Psi^s(\cdot)$ for $s \in \{0, \cdots, S\}$ can be solved as

$$\Psi^s(z) = \frac{\hat{\Psi}^s(q_0(z))}{\sum_{s\in\{0,\cdots,S\}} \hat{\Psi}^s(q_0(z))}, \quad \text{where} \quad \hat{\Psi}^s(q_0) \equiv \sum_{a\in\mathcal{A}} \Pr(a)\hat{\Psi}^s_a(q_0).$$

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

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

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

$$\begin{aligned}
\hat{\Psi}^1_a(q_0) &= \mu \int_0^{t(q_0,\tilde{q})} \theta^1_a e^{-(\mu+\theta^1_a)x} \left[e^{-(\mu+\theta^2_a)t(\hat{q}(q(s)),\overline{q}_2)} \int_0^\infty e^{\mu\tau} d\tau \right] dx + \mu \int_{t(q_0,\tilde{q})}^{t(q_0,\overline{q}_1)} \theta^1_a e^{-(\mu+\theta^1_a)x} \left[\int_0^\infty e^{\mu\tau} d\tau \right] dx \\
&= \theta^1_a \left[\frac{\theta^1_0}{\theta^1_1} \frac{(1-q_0)\overline{q}_2}{(1-\overline{q}_2)q_0} \right]^{\frac{\mu+\theta^2_a}{\Delta\theta^2}} \frac{1 - e^{-\left[\mu+\theta^1_a - \frac{\Delta\theta^1_a}{\Delta\theta^2_a}(\mu+\theta^2_a)\right]t(q_0,\tilde{q})}}{\mu+\theta^1_i - \frac{\Delta\theta^1_a}{\Delta\theta^2_a}(\mu+\theta^2_a)} + \frac{\theta^1_a}{\mu+\theta^1_a} \left[e^{-(\mu+\theta^1_a)t(q_0,\tilde{q})} - e^{-(\mu+\theta^1_a)t(q_0,\overline{q}_1)} \right],
\end{aligned}$$

where $\tilde{q} = \max\{\hat{q}^{-1}(\overline{q}_2), \overline{q}_1\}$. 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 beliefs fall below $\overline{q}_2$ become some college attendees.

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

$$\begin{aligned}
\hat{\Psi}^2_a(q_0) &= \theta^2_a \int_0^\infty e^{-\mu\tau} d\tau \left\{ \mu \int_0^{t(q_0,\tilde{q})} \theta^1_a e^{-(\mu+\theta^1_a)x} \left[\int_0^{t(\hat{q}(q(x)),\overline{q}_2)} e^{-(\mu+\theta^2_a)\tau} d\tau \right] dx \right\} \\
&= \frac{\theta^1_a\theta^2_a}{(\mu+\theta^2_a)(\mu+\theta^1_a)} \left[1 - e^{-(\mu+\theta^1_a)t(q_0,\tilde{q})} \right] - \frac{\theta^1_a\theta^2_a \left(\frac{\theta^1_0}{\theta^1_1} \frac{(1-q_0)\overline{q}_2}{(1-\overline{q}_2)q_0} \right)^{\frac{\mu+\theta^2_a}{\Delta\theta^2}}}{(\mu+\theta^2_a)} \left[\frac{1 - e^{-\left(\mu+\theta^1_a - \frac{\Delta\theta^1_a}{\Delta\theta^2_a}(\mu+\theta^2_a)\right)t(q_0,\tilde{q})}}{\mu+\theta^1_a - \frac{\Delta\theta^1_a}{\Delta\theta^2_a}(\mu+\theta^2_a)} \right].
\end{aligned}$$

E Numerical Method

This section describes the numerical procedure for calculating the stationary equilibrium. It relies on the analytical results derived in Section 4.1, which considerably reduce the computational burden. We also describe how to find a global optimum in the calibration.

Stationary Equilibrium. The algorithm to obtain a stationary equilibrium is the following.

Step 1: Guess a wage distribution G.

Step 2: Obtain an initial skill distribution $Z_a(\cdot; G)$ using equation (1).

Step 3: Calculate the belief cutoffs $\overline{q}^s$ that deliver the distribution of educational attainments observed in the data using the expressions in Appendix D.

Step 4: Choose an initial guess for returns to education R.

Step 5: Solve the labor market stage.

Step 6: Calculate the value of being unemployed U^s using equation (18). Calculate the implicit returns to education R' that rationalizes $\overline{q}^s$ in equation (21).

Step 7: If $||R' - R|| > \varepsilon$ for some small ε, set $R = R'$ and return to Step 5.

Step 8: Simulate the model to generate a wage distribution G'.

Step 9: Stop if G' is sufficiently close to G. Otherwise, set $G = G'$ and return to Step 2.

In Step 8, $100,000$ wages are drawn to construct a new wage distribution. In Step 9, we evaluate the convergence of the distribution by its 1st to 4th moments (i.e., mean, variance, skewness, kurtosis). We have checked that the system behaves as a contraction, and the equilibrium thus seems unique.

Global Optimization Algorithm. The model is highly nonlinear. To ensure that our solution ϑ^* to problem (26) is a global optimum, we use the following global optimization 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 $\tilde{\vartheta}_i$.

Step 3: Construct a quasi-random parameter vector $\hat{\vartheta}_i \in \Theta$ generated by a Sobol sequence.

Step 4: Set the new initial guess by $\vartheta_i^{new} = f(i)\tilde{\vartheta}_i^* + (1 - f(i))\hat{\vartheta}_i$ where $\tilde{\vartheta}_i^*$ is the best solution obtained up until iteration i and $f : \mathbb{N} \to [0, 1]$ is an increasing weighting function.

Step 5: Set $i = i + 1$, $\vartheta_i = \vartheta_{i-1}^{new}$ and go back to Step 2.

Step 6: Stop if converged.

Calibration Table A1 presents the calibrated parameters and the targeted moments.

Table A1: Calibrated Parameters and Targeted Moments

Parameters	1970	2015	Targeted Moments	Data 1970	Data 2015	Model 1970	Model 2015
Exactly Identified							
R_s	$\begin{bmatrix}1\\1.03\\1.20\end{bmatrix}$	$\begin{bmatrix}1\\1.03\\1.21\end{bmatrix}$	Educational attainment distribution (HS,SC,CL)	$\begin{bmatrix}.544\\.194\\.262\end{bmatrix}$	$\begin{bmatrix}.325\\.289\\.386\end{bmatrix}$	$\begin{bmatrix}.544\\.194\\.262\end{bmatrix}$	$\begin{bmatrix}.325\\.289\\.386\end{bmatrix}$
δ_s	$\begin{bmatrix}.006\\.010\\.002\end{bmatrix}$	$\begin{bmatrix}.013\\.007\\.001\end{bmatrix}$	Unemployment rate (%)	$\begin{bmatrix}2.3\\2.2\\1.1\end{bmatrix}$	$\begin{bmatrix}6.5\\4.7\\2.4\end{bmatrix}$	$\begin{bmatrix}2.3\\2.2\\1.1\end{bmatrix}$	$\begin{bmatrix}6.5\\4.7\\2.4\end{bmatrix}$
χ	.0003	.0108	Unemp. duration (months)	2.82	6.49	2.82	6.49
Calibrated							
μ_a	$(-.072, -.065)$		Between-Group: SC-HS	15.2	22.4	8.8	19.9
σ_a	$(.36, .38)$		Between-Group: CL-HS	39.0	74.2	38.5	73.3
α_a	$(.37, .39)$		Within-Group: HS	0.14	0.29	0.14	0.27
γ	31.7	9.9	Within-Group: SC	0.18	0.30	0.18	0.32
β	0.27		Within-Group: CL	0.25	0.39	0.25	0.38

Note: The parameter μ_{a_0} is set for normalization. Between-group inequality is measured as the difference in average log earnings for two educational groups. Within-group inequality is defined as the conditional variance of log earnings. Between-group inequality of SC relative to HS is not targeted.

F Supplemental Results

Figure A2 is a supplemental figure to Figure 3. Panel A shows that a lower α_{a_0} works in the same way as a higher α_{a_1}, and hence, what truly matters is the difference $\alpha_{a_1} - \alpha_{a_0}$. Figure A2 shows that we obtain the same patterns as Figure 3 for different schooling levels (Panels B, E and F) or points in time (Panels C and D).

Figure A3 is a supplemental figure to Figure 4. It plots the distributions of firm size and output relative to their means in 1970. The model does a good job in replicating the overall shape of these distributions.

Table A2 displays the results of our counterfactual analysis and decomposition for SC. The trends in the between- and within-group inequality measures behave like they

Table A2: Counterfactual Analysis and Decomposition: SC

	[1] **1970**	[2] Ψ_{70}	[3] Z_{70}	[4] **2015**
Between-group: SC-HS	**11.1**	**19.8**	**7.8**	**11.5**
Returns to Education	2.7	2.7	2.7	2.8
Worker Skill Difference	5.0	5.0	2.7	6.2
Firm Pay Difference	3.4	12.1	2.5	2.5
B. Within-group SC	**0.18**	**0.33**	**0.29**	**0.33**
Person effect	0.17	0.17	0.15	0.18
Firm effect	0.01	0.08	0.08	0.08
Friction effect	0.00	0.08	0.07	0.07

Note: The inequality measures in 1970 and 2015 are different from those in Table 1 as the time-variant labor-market parameters (χ and δ) are set at the average levels between 1970 and 2015.

do for CL and HS, while there is little increase in between-group inequality from 1970 to 2015. Returns to education increase from 2.7 log points in 1970 to 2.8 log points in 2015. The worker skill difference increases from 5.0 log points to 6.2 log points, while the firm pay difference decreases from 3.4 log points to 2.5 log points. Most of the increase in within-group inequality is accounted for by the firm and friction effects.

Column [2] shows that between-group inequality increases from 11.1 log points in column [1] to 19.8 log points in column [2], but column [3] shows that allowing for education choices reduces it to 7.8 log points. This means that like the case for CL, we can see the attenuation of the increasing dispersion effect via endogenous education decisions and the amplification via intergenerational linkages. However, in contrast to the case for CL, the attenuation effect is stronger for SC; when we allow for the skill supply responses, moving from column [2] to [4], between-group inequality in fact decreases by 8.3 log points.

References

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.

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

Hartman-Glaser, B., H. N. Lustig, and M. Z. Xiaolan (2019). Capital share dynamics when firms insure workers. *Journal of Finance 74*(4), 1707–1751.

Jaeger, D. A. and M. E. Page (1996). Degrees matter: New evidence on sheepskin effects in the returns to education. *The Review of Economics and Statistics 78*(4), 733–740.

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.

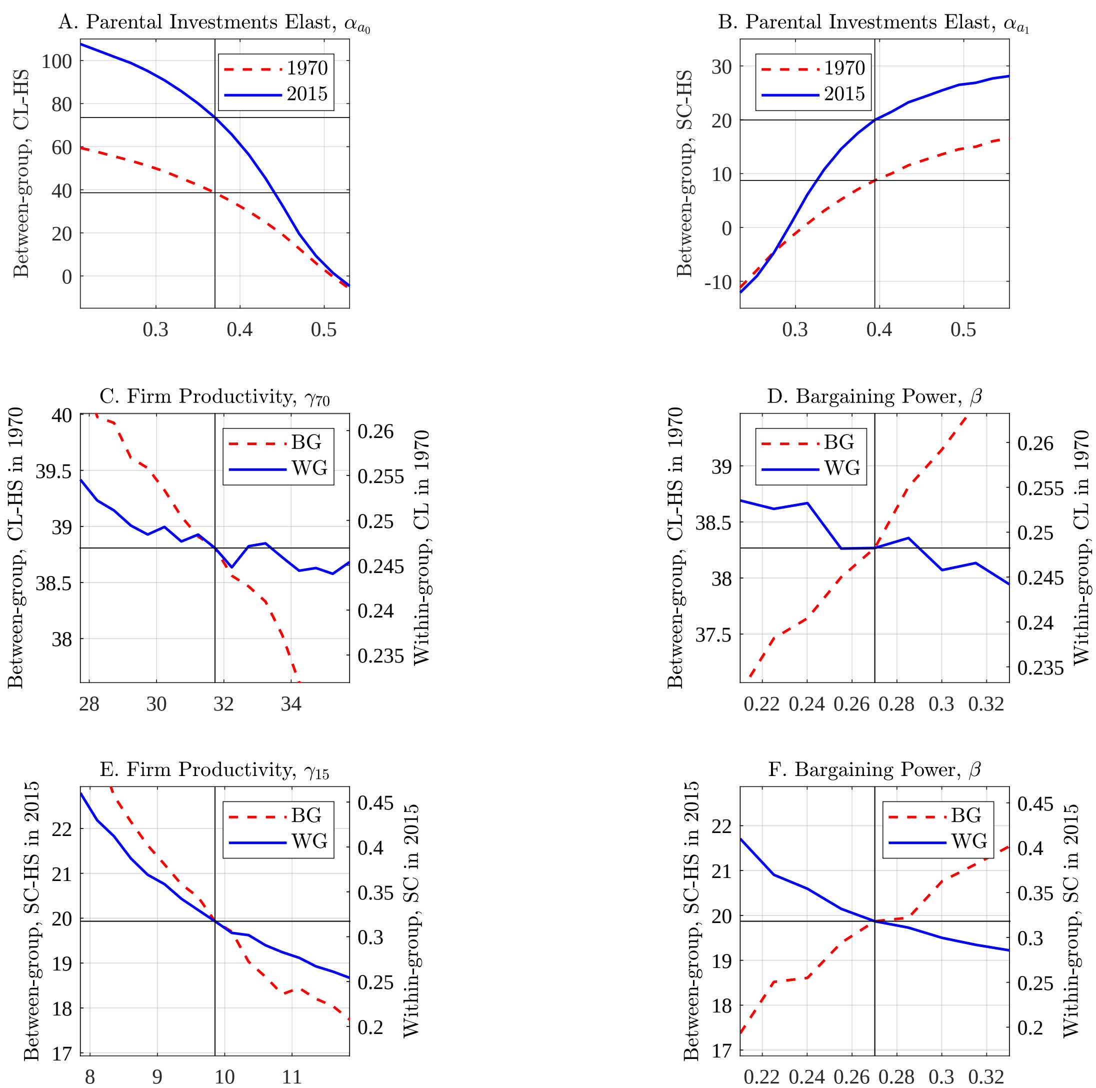

Figure A2: Identification (supplement). Between-group inequality in 1970 and 2015 is plotted against α_{a_0} for CL (Panel A) and against α_{a_1} for SC (Panel B). Between- and within-group inequality measures against γ and β are plotted for CL in 1970 (Panels C and D) and for SC in 2015 (Panels E and F). The vertical and horizontal lines correspond to the calibrated parameter value and the value of the model moment, respectively.

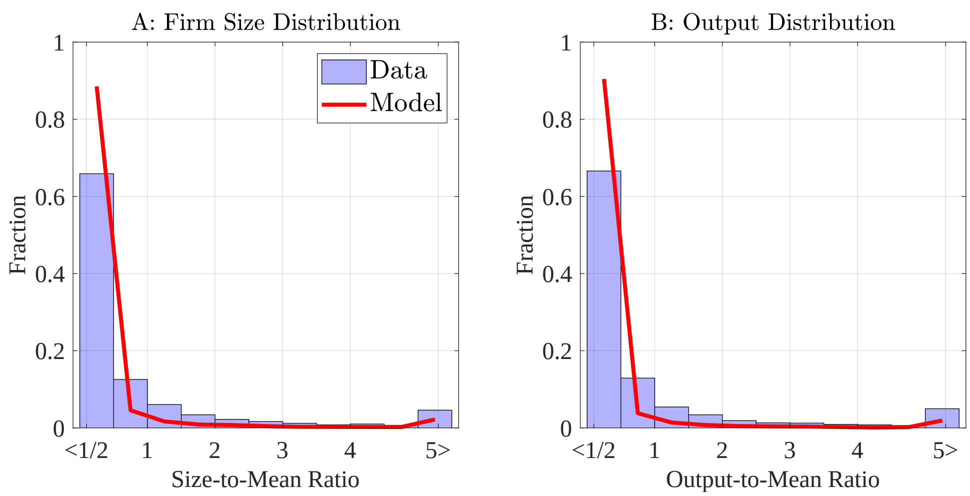

Figure A3: Firm Size and Output Distribution. The figure plots the distributions of firm size (Panel A) and output (Panel B) in the data and the model in 1970. The values are normalized relative to mean; e.g., the leftmost bar in Panel A shows that the fraction of firms whose size is below half of the mean size.